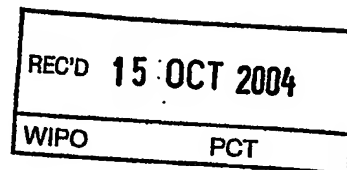




PCT/FR 2004/001800



# BREVET D'INVENTION

**CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION**

## COPIE OFFICIELLE

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le 09 JUL. 2004

Pour le Directeur général de l'Institut  
national de la propriété industrielle  
Le Chef du Département des brevets

Martine PLANCHE

**DOCUMENT DE PRIORITÉ**

PRÉSENTÉ OU TRANSMIS  
CONFORMÉMENT À LA  
RÈGLE 17.1.a) OU b)

**BEST AVAILABLE COPY**

INSTITUT  
NATIONAL DE  
LA PROPRIÉTÉ  
INDUSTRIELLE

SIEGE  
26 bis, rue de Saint-Petersbourg  
75800 PARIS cedex 08  
Téléphone : 33 (0)1 53 04 53 04  
Télécopie : 33 (0)1 53 04 45 23  
www.inpi.fr

<b>REMISE DES PIÈCES</b> DATE <b>8 JUIL 2003</b> LIEU <b>75 INPI PARIS</b> N° D'ENREGISTREMENT <b>0308350</b> NATIONAL ATTRIBUÉ PAR L'INPI DATE DE DÉPÔT ATTRIBUÉE <b>- 8 JUIL. 2003</b> PAR L'INPI		<b>1</b> NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE À QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE  <b>GROSSET-FOURNIER &amp; DEMACHY</b> 54, rue Saint-Lazare F-75009 Paris	
<b>Vos références pour ce dossier</b> (facultatif) <b>IFB 03 BP CNR GFAT</b>			
<b>Confirmation d'un dépôt par télécopie</b> <input type="checkbox"/> N° attribué par l'INPI à la télécopie			
<b>2 NATURE DE LA DEMANDE</b>		<b>Cochez l'une des 4 cases suivantes</b>	
Demande de brevet		<input checked="" type="checkbox"/>	
Demande de certificat d'utilité		<input type="checkbox"/>	
Demande divisionnaire		<input type="checkbox"/>	
<i>Demande de brevet initiale</i> <i>ou demande de certificat d'utilité initiale</i>		N° _____ Date ____ / ____ / ____ N° _____ Date ____ / ____ / ____	
Transformation d'une demande de brevet européen <i>Demande de brevet initiale</i>		<input type="checkbox"/> N° _____ Date ____ / ____ / ____	
<b>3 TITRE DE L'INVENTION (200 caractères ou espaces maximum)</b> GLUTAMINE:FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE (GFAT) COMPRENANT UNE ETIQUETTE DE PURIFICATION INTERNE, ET SON UTILISATION POUR LE CRIBLAGE DE COMPOSES			
<b>4 DÉCLARATION DE PRIORITÉ</b> OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE FRANÇAISE		Pays ou organisation _____ N° _____ Date ____ / ____ / ____ Pays ou organisation _____ N° _____ Date ____ / ____ / ____ Pays ou organisation _____ N° _____ Date ____ / ____ / ____ <input type="checkbox"/> S'il y a d'autres priorités, cochez la case et utilisez l'imprimé «Suite»	
<b>5 DEMANDEUR</b>		<input type="checkbox"/> S'il y a d'autres demandeurs, cochez la case et utilisez l'imprimé «Suite»	
Nom ou dénomination sociale		<b>CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE</b>	
Prénoms			
Forme juridique			
N° SIREN			
Code APE-NAF			
Adresse	Rue	<b>3, rue Michel-Ange</b>	
	Code postal et ville		
Pays	<b>F-75794 PARIS CEDEX 16</b>		
Nationalité	<b>FRANCE</b>		
N° de téléphone (facultatif)	<b>FRANCAISE</b>		
N° de télécopie (facultatif)			
Adresse électronique (facultatif)			

**BREVET D'INVENTION  
CERTIFICAT D'UTILITÉ**

REQUÊTE EN DÉLIVRANCE 2/2

Réservé à l'INPI

REMISE DES PIÈCES

DATE

8 JUIL 2003

LIEU

75 INPI PARIS

N° D'ENREGISTREMENT

0308350

NATIONAL ATTRIBUÉ PAR L'INPI

DB 540 W /260899

<b>Vos références pour ce dossier :</b> (facultatif)		<b>IFB 03 BP CNR GFAT</b>	
<b>6 MANDATAIRE</b>			
Nom		<b>DEMACHY</b>	
Prénom		<b>Charles</b>	
Cabinet ou Société		<b>GROSSET-FOURNIER &amp; DEMACHY</b>	
N° de pouvoir permanent et/ou de lien contractuel			
Adresse	Rue	<b>54, rue Saint-Lazare</b>	
	Code postal et ville	<b>75009 PARIS</b>	
N° de téléphone (facultatif)		<b>01.42.81.09.58</b>	
N° de télécopie (facultatif)		<b>01.42.81.08.71</b>	
Adresse électronique (facultatif)			
<b>7 INVENTEUR (S)</b>			
Les inventeurs sont les demandeurs		<input type="checkbox"/> Oui <input checked="" type="checkbox"/> Non Dans ce cas fournir une désignation d'inventeur(s) séparée	
<b>8 RAPPORT DE RECHERCHE</b>		Uniquement pour une demande de brevet (y compris division et transformation)	
Établissement immédiat ou établissement différé		<input type="checkbox"/> <input type="checkbox"/>	
Paiement échelonné de la redevance		<b>Paiement en deux versements, uniquement pour les personnes physiques</b> <input checked="" type="checkbox"/> Oui <input type="checkbox"/> Non	
<b>9 RÉDUCTION DU TAUX DES REDEVANCES</b>		<b>Uniquement pour les personnes physiques</b> <input type="checkbox"/> Requête pour la première fois pour cette invention (joindre un avis de non-imposition) <input type="checkbox"/> Requête antérieurement à ce dépôt (joindre une copie de la décision d'admission pour cette invention ou indiquer sa référence) :	
Si vous avez utilisé l'imprimé «Suite», indiquez le nombre de pages jointes			
<b>10 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE</b> (Nom et qualité du signataire)		<b>Charles DEMACHY</b> Mandataire 422.5/PP.170	<b>VISA DE LA PRÉFECTURE OU DE L'INPI</b>  <b>L. MARIELLO</b>

# GLUTAMINE:FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE (GFAT) COMPRENANT UNE ETIQUETTE DE PURIFICATION INTERNE, ET SON UTILISATION POUR LE CRIBLAGE DE COMPOSES

La présente invention concerne une glutamine:fructose-6-phosphate amidotransférase modifiée, purifiable rapidement et en quantités suffisantes pour le criblage de composés modifiant son activité.

Les glutamine:fructose-6-phosphate amidotransférases (GFAT), EC 2.6.1.16, également appelées glucosamine-6-phosphate synthases ou 2-déoxy-glucose-6-phosphate céto isomérases, sont impliquées dans la voie de biosynthèse des hexosamines. La GFAT catalyse la première étape, limitante, de cette voie de biosynthèse selon la réaction :

$$\text{L-Glutamine} + \text{fructose-6-phosphate} \longrightarrow \text{L-Glutamate} + \text{glucosamine-6-phosphate}$$
 par transfert de l'azote amidique de la L-Glutamine sur la fonction cétone du fructose-6-phosphate. Les GFAT contrôlent donc le flux de glucose dans la voie des hexosamines, via le fructose-6-phosphate, et par conséquent la formation des hexosamines produites.

Une forme bactérienne recombinante de la GFAT, la glucosamine-6-phosphate synthase d'*Escherichia coli*, a été purifiée à homogénéité et étudiée de façon exhaustive. Les propriétés et le mécanisme enzymatique du transfert de l'amide ont notamment été largement décrits (pour revue Teplyakov *et al.*, *Nat. Prod. Rep.* (2002) 19:60). En particulier, cette enzyme, dont la structure cristalline a été résolue (Teplyakov *et al.*, *J. Mol. Biol.* (2001) 313:1093), est formée de deux domaines, l'un ayant une activité hydrolase (glutaminase) et l'autre une activité isomérase.

Par ailleurs, des GFAT d'eucaryotes ont été caractérisées, dont notamment celle de foie de rat (Huynh *et al.*, *Arch. Biochem. Biophys.* (2000) 379:307) et celle de la levure *Candida albicans* (Milewsky *et al.*, *J. Biol. Chem.* (1999) 274:4000).

Chez l'homme, des études préliminaires ont montré la présence d'une activité GFAT dans le foie (Ghosh *et al.*, *J. Biol. Chem.* (1960) 235:1265). Plusieurs GFAT sont désormais connues. GFAT1, la forme principale, GFAT2, qui est exprimée préférentiellement dans le système nerveux central, et GFAT1Alt, une isoforme de GFAT1, exprimée essentiellement dans les muscles striés. Les séquences peptidiques de GFAT1 et GFAT2 possèdent 75% d'identités de séquences entre elles, et celles de GFAT1 et GFAT1Alt sont identiques excepté pour une insertion de 18 acides aminés dans la séquence de GFAT1Alt. Les séquences de GFAT sont donc très conservées chez

l'homme, mais également entre les espèces, puisque les séquences peptidiques de la GFAT1 humaine et de la GFAT d'*E. coli* ou de la GFAT1 de souris présentent respectivement 35% et 99% d'identités.

Le gène de la GFAT1 humaine a été cloné en 1992 (McKnight *et al.*, *J. Biol. Chem.* (1992) 267:25208). Il code une protéine de 77 kDa formée de deux domaines distincts (Teplyakov *et al.*, *Nat. Prod. Rep.* (2002) 19:60).

L'augmentation de la production de l'UDP-NAc-GlcNH<sub>2</sub>, le produit final de la voie de biosynthèse des hexosamines, et son accumulation dans les tissus ont récemment été impliquées dans le développement de la résistance à l'insuline (Marshall *et al.*, *FASEB J.* (1991) 5:3031, Yki-Jarvinen *et al.*, *Diabetes* (1996) 45:302, Thompson *et al.*, *J. Biol. Chem.* (1997) 272: 7759, Hawkins *et al.*, *J. Clin. Invest.* (1997) 99:2173, Robinson *et al.*, *Diabetes* (1993) 42:1333, Daniels *et al.*, *J. Clin. Invest.* (1995) 96:1235, Baron *et al.*, *J. Clin. Invest.* (1995) 96:2792).

Ainsi, il a été montré qu'une augmentation du niveau cellulaire d'UDP-NAc-GlcNH<sub>2</sub> par une modeste surexpression de GFAT1, ou un apport de glucosamine exogène, peut induire une résistance à l'insuline à la fois in vivo et dans des adipocytes en culture (Robinson *et al.*, *Diabetes* (1993) 42:1333, Daniels *et al.*, *J. Clin. Invest.* (1995) 96:1235, Baron *et al.*, *J. Clin. Invest.* (1995) 96:2792, Hebert *et al.*, *J. Clin. Invest.* (1996) 98:930).

En effet, l'insuline active sa voie de transduction en se fixant à son récepteur, ce qui induit la translocation des transporteurs de glucose, tels que le récepteur-GLUT4, stockés dans la cellule, vers la membrane, et augmente l'afflux de glucose. Le glucose entre ainsi dans la voie de la glycolyse et est converti en glucose-6-phosphate puis en fructose-6-phosphate. Lorsque l'afflux de glucose est excessif, le fructose-6-phosphate entre dans la voie de biosynthèse des hexosamines et est converti en glucosamine-6-phosphate par la GFAT. Plusieurs observations indiquent que les métabolites de la glucosamine-6-phosphate empêchent la translocation des récepteurs au glucose vers la membrane cellulaire, diminuant ainsi l'afflux du glucose cellulaire (Marshall *et al.*, *FASEB J.* (1991) 5:3031, Giacarri *et al.*, *Diabetologia* (1995) 38:518, Marshall *et al.*, *J. Biol. Chem.* (1991) 266:4706, Paterson *et al.*, *Endocrinology* (1995) 136:2809).

Le mécanisme par lequel les métabolites de la glucosamine-6-phosphate exercent leurs effets physiologiques n'est pas clair. Une hypothèse a cependant été proposée : une concentration cytosolique élevée d'UDP-NAc-GlcNH<sub>2</sub> entraînerait l'hyperglycosylation des sites de phosphorylation Ser ou Thr, conduisant de la sorte à

l'arrêt de la voie de signalisation de l'insuline (Comer *et al.*, *J. Biol. Chem.* (2000) 275:29179).

L'activité GFAT est donc considérée comme étant l'une des causes des hauts niveaux de glucose sanguin ; par ailleurs elle est connue pour être élevée chez les patients atteints de diabète sucré non-insulino dépendant ou diabète de type II (Yki-Jarvinen *et al.*, *Diabetes* (1996) 45:302).

L'obtention d'inhibiteurs de la GFAT permettrait d'abaisser la glycémie en particulier chez les individus atteints de pathologies liées à une hyperglycémie, telles que le diabète de type II, l'acidose et/ou la cétose diabétique, par exemple.

Des inhibiteurs de la GFAT de plante ou de champignon pourraient également permettre d'obtenir respectivement des fongicides et des herbicides.

Toutefois, malgré l'obtention de formes recombinantes de GFAT, l'instabilité des préparations enzymatiques obtenues, leur faible quantité, et leur niveau de purification insuffisant, n'ont pas permis d'obtenir des inhibiteurs efficaces de GFAT.

Un objet de l'invention est donc de fournir une GFAT modifiée dont l'activité est stable et pouvant être obtenue en grande quantité, avec un haut niveau de pureté et d'activité.

La présente invention concerne une protéine enzymatiquement active comprenant :

- une séquence de GFAT et au moins une séquence d'une étiquette de purification, la séquence de l'étiquette de purification étant insérée entre deux acides aminés consécutifs de la séquence de GFAT, ou
- une séquence dérivant de la séquence précédente par suppression, insertion ou mutation d'au moins un acide aminé, sous réserve que ladite protéine présente une activité enzymatique, ou
- une séquence présentant au moins 35 %, notamment au moins 90 %, d'identité de séquence et/ou au moins 44 %, notamment au moins 95 %, de similarité de séquence avec l'une des séquences précédentes, sous réserve que ladite protéine présente une activité enzymatique.

On désigne par GFAT une enzyme de la classe E.C. 2.6.1.16 catalysant la réaction suivante :

L-Glutamine + fructose-6-phosphate  $\longrightarrow$  L-Glutamate + glucosamine-6-phosphate  
notamment dans les conditions expérimentales telles qu'elles sont décrites dans l'exemple qui suit ou dans Broschat *et al.*, *J. Biol. Chem.* (2002) 277:14764.

GFAT est désignée sous le nom de glutamine:fructose-6-phosphate amidotransférase, ou également glucosamine-6-phosphate synthase ou 2-déoxy-glucose-6-phosphate cétoI isomérase.

On désigne par « protéine enzymatiquement active » une protéine ayant une action catalytique.

Avantageusement, la protéine enzymatiquement active possède une activité GFAT.

On désigne par « étiquette de purification » une séquence peptidique susceptible de se lier spécifiquement à un ligand donné. Avantageusement la liaison dudit ligand à l'étiquette de purification permet de former un complexe entre la protéine portant l'étiquette de purification et ledit ligand, ledit complexe pouvant être spécifiquement isolé.

Avantageusement les étiquettes de purification selon l'invention ne sont pas placées en bout de chaîne peptidique, à l'extrémité N-terminale ou C-terminale, mais à l'intérieur de la chaîne peptidique.

On désigne par « identité de séquence » le pourcentage d'acides aminés identiques entre deux séquences alignées, notamment à l'aide d'algorithmes tels que celui défini par Altschul *et al.*, *Nucleic Acids Res.* (1997) 25:3389, par exemple.

On désigne par « similarité de séquence » le pourcentage d'acides aminés similaires, c'est-à-dire d'acides aminés dont les chaînes latérales possèdent des propriétés physico-chimiques proches, entre deux séquences alignées, notamment à l'aide d'algorithmes tels que celui défini par Altschul *et al.*, *Nucleic Acids Res.* (1997) 25:3389, par exemple.

La présente invention concerne en particulier une protéine telle que définie ci-dessus, dans laquelle la séquence de GFAT correspond à une séquence de GFAT de bactérie ou d'eucaryote, notamment de plante, de champignon ou d'animal, en particulier d'insecte ou de mammifère, plus particulièrement d'homme.

L'invention concerne plus particulièrement une protéine telle que définie ci-dessus, dans laquelle la séquence de l'étiquette de purification est insérée entre deux acides aminés consécutifs de la séquence de GFAT, lesdits acides aminés étant compris dans :

- une partie de la séquence de la GFAT correspondant à la séquence s'étendant entre le feuillet  $\beta 2$  et le feuillet  $\beta 3$  de la GFAT d'*Escherichia coli*, ou

- une partie de la séquence de la GFAT correspondant à la séquence s'étendant entre le feuillet  $\beta 13$  et le feuillet  $\beta 14$  de la GFAT d'*Escherichia coli*, ou
- une partie de la séquence de la GFAT correspondant à la séquence s'étendant entre le feuillet  $\beta 15$  et l'hélice  $\alpha 6$  de la GFAT d'*Escherichia coli*.

La structure de la GFAT d'*Escherichia coli* est décrite en particulier par Teplyakov *et al.*, *J. Mol. Biol.* (2001) 313:1093 (protéine entière), par Isupov *et al.*, *Structure* (1996) 4:801 (domaine glutaminase) et par Teplyakov *et al.*, *Structure* (1998) 6:1047 (domaine isomérase). La structure de la protéine complète est notamment consultable à l'aide du fichier de coordonnées atomiques 1JXA déposé auprès de la Protein Data Bank (<http://www.pdb.org>).

La séquence peptidique de la GFAT d'*E. coli* est définie par SEQ ID NO : 13.

La séquence s'étendant entre le feuillet  $\beta 2$  et le feuillet  $\beta 3$  correspond à la séquence s'étendant approximativement entre les acides aminés 30 à 80 de la GFAT d'*E. coli*, situés dans le domaine glutaminase.

La séquence s'étendant entre le feuillet  $\beta 13$  et le feuillet  $\beta 14$  correspond à la séquence s'étendant approximativement entre les acides aminés 220 à 230 de la GFAT d'*E. coli*, situés dans le domaine glutaminase.

La séquence s'étendant entre le feuillet  $\beta 15$  et l'hélice  $\alpha 6$  correspond à la séquence s'étendant approximativement entre les acides aminés 235 à 250 de la GFAT d'*E. coli*, situés dans le domaine isomérase.

L'identification des parties de séquences d'une GFAT correspondant à des structures secondaires de la GFAT d'*E. coli* peut être obtenue en alignant la séquence de ladite GFAT avec celle de la GFAT d'*E. coli*, notamment à l'aide d'un algorithme tel que celui défini par Altschul *et al.*, *Nucleic Acids Res.* (1997) 25:3389, par exemple.

L'invention concerne notamment une protéine ci-dessus, dans laquelle la séquence de GFAT correspond à :

- SEQ ID NO : 2, correspondant à la séquence de la GFAT1 humaine,
- SEQ ID NO : 4, correspondant à la séquence de la GFAT2 humaine,
- SEQ ID NO : 6, correspondant à la séquence de la GFAT1Alt humaine.

La séquence de la GFAT1 humaine est notamment décrite dans McKnight *et al.*, *J. Biol. Chem.* (1992) 267:25208, et correspond à la séquence nucléotidique SEQ ID NO : 1.

La séquence de la GFAT2 humaine est notamment décrite dans Oki *et al.*, *Genomics* (1999) 57:227, et correspond à la séquence nucléotidique SEQ ID NO : 3.



La séquence de la GFAT1Alt humaine est notamment décrite dans DeHaven *et al.*, *Diabetes* (2001) 50:2419, et correspond à la séquence nucléotidique SEQ ID NO : 5.

L'invention concerne en particulier une protéine ci-dessus, dans laquelle la séquence de l'étiquette de purification est insérée entre deux acides aminés consécutifs, lesdits acides aminés étant compris entre les acides aminés :

- 43 à 47, 298 à 306, et/ou 342 à 347 de SEQ ID NO : 2,
- 42 à 45, 299 à 307, et/ou 343 à 348 de SEQ ID NO : 4
- 43 à 47, 316 à 324, et/ou 360 à 365 de SEQ ID NO : 6

Les acides aminés 43 à 47 de SEQ ID NO : 2, 42 à 45 de SEQ ID NO : 4 et 43 à 47 de SEQ ID NO : 6 correspondent à la partie de la séquence de la GFAT d'*E. coli* s'étendant entre le feuillet  $\beta 2$  et le feuillet  $\beta 3$ .

Les acides aminés 298 à 306 de SEQ ID NO : 2, 299 à 307 de SEQ ID NO : 4 et 325 à 330 de SEQ ID NO : 6 correspondent à la partie de la séquence de la GFAT d'*E. coli* s'étendant entre le feuillet  $\beta 13$  et le feuillet  $\beta 14$ .

Les acides aminés 342 à 347 de SEQ ID NO : 2, 343 à 348 de SEQ ID NO : 4 et 360 à 365 de SEQ ID NO : 6 correspondent à la partie de la séquence de la GFAT d'*E. coli* s'étendant entre le feuillet  $\beta 15$  et l'hélice  $\alpha 6$ .

L'invention concerne plus particulièrement une protéine ci-dessus, dans laquelle la séquence de l'étiquette de purification est insérée entre les acides aminés :

- 299 et 300 de SEQ ID NO : 2,
- 300 et 301 de SEQ ID NO : 4,
- 317 et 318 de SEQ ID NO : 6.

L'invention concerne notamment une protéine ci-dessus, dans laquelle l'étiquette de purification correspond à une séquence d'environ 2 à environ 10 acides aminés, notamment d'environ 4 à environ 8 acides aminés.

Des étiquettes de purification préférées selon l'invention concernent notamment des étiquettes dites FLAG (Sigma-Aldrich, France). Ces étiquettes se lient spécifiquement à un paratope donné, ledit paratope pouvant appartenir à un anticorps ou à un fragment d'anticorps par exemple. Un exemple particulier d'étiquette FLAG est constitué par la séquence peptidique Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys (SEQ ID NO : 18) par exemple.

D'autres étiquettes préférées selon l'invention sont des étiquettes formées de plusieurs histidines. Ces étiquettes peuvent former des complexes avec des cations métalliques divalents tels que  $\text{Ni}^{2+}$  ou  $\text{Co}^{2+}$  par exemple.

L'invention concerne en particulier une protéine telle que définie ci-dessus, dans laquelle l'étiquette de purification est une hexa-histidine.

On désigne par hexa-histidine la séquence His-His-His-His-His-His (SEQ ID NO : 19).

5 L'invention concerne plus particulièrement une protéine telle que définie ci-dessus correspondant aux séquences :

- SEQ ID NO : 8, correspondant à la séquence SEQ ID NO : 2 dans laquelle une hexa-histidine est insérée entre les acides aminés 299 et 300,
- SEQ ID NO : 10, correspondant à la séquence SEQ ID NO : 4 dans laquelle une
- 10 hexa-histidine est insérée entre les acides aminés 300 et 301, et
- SEQ ID NO : 12, correspondant à la séquence SEQ ID NO : 6 dans laquelle une hexa-histidine est insérée entre les acides aminés 317 et 318.

La présente invention concerne également un acide nucléique comprenant ou étant constitué d'une séquence codant pour une protéine telle que définie ci-dessus.

15 L'invention concerne plus particulièrement un acide nucléique comprenant ou étant constitué de la séquence nucléotidique :

- SEQ ID NO : 7 codant pour la protéine SEQ ID NO : 8, ou
- SEQ ID NO : 9 codant pour la protéine SEQ ID NO : 10, ou
- SEQ ID NO : 11 codant pour la protéine SEQ ID NO : 12,

20 ou de son complémentaire, ou étant dérivé de ladite séquence par mutation, insertion ou délétion d'au moins un nucléotide, sous réserve que ladite séquence nucléotidique code pour une protéine enzymatiquement active.

Selon un autre mode de réalisation, la présente invention concerne également un vecteur eucaryote ou procaryote comprenant un acide nucléique tel que défini ci-dessus.

25 Ces vecteurs permettent notamment de synthétiser les protéines selon l'invention dans un organisme eucaryote ou procaryote.

Avantageusement l'invention concerne un vecteur d'expression de type baculovirus permettant la synthèse des protéines selon l'invention dans des cellules d'insecte.

30 La présente invention concerne également un procédé de purification d'une protéine telle que définie ci-dessus, à partir d'une solution comprenant ladite protéine, comprenant une étape de mise en présence de ladite solution avec un composé se liant spécifiquement à l'étiquette de purification de ladite protéine et une étape de séparation

du complexe formé par la liaison de ladite protéine audit composé des autres constituants de la solution.

Le composé peut être fixé sur un support solide de sorte que le complexe formé entre ledit composé et ladite protéine peut être récupéré par centrifugation ou filtration. Optionnellement ledit composé fixé sur son support peut être disposé dans une colonne à travers laquelle ladite solution est éluée.

Avantageusement, le procédé ci-dessus peut également comprendre une étape de dissociation du complexe formé par la liaison de ladite protéine audit composé afin de récupérer la protéine purifiée.

L'invention concerne plus particulièrement un procédé de purification tel que défini ci-dessus, comprenant une étape de mise en présence d'une solution comprenant une protéine telle que définie ci-dessus avec un composé comprenant un cation métallique divalent tel que  $\text{Ni}^{2+}$  ou  $\text{Co}^{2+}$ , notamment  $\text{Ni}^{2+}$ , et une étape de séparation du complexe formé par la liaison de la protéine audit composé des autres constituants de la solution.

Avantageusement, le procédé ci-dessus peut également comprendre une étape de dissociation du complexe formé par la liaison de ladite protéine audit composé comprenant un cation métallique divalent, notamment à l'aide d'imidazole, afin de récupérer la protéine purifiée.

Selon un autre mode de réalisation la présente invention concerne un procédé de conservation d'une protéine telle que définie ci-dessus sous une forme enzymatiquement active, notamment à  $-80^{\circ}\text{C}$  ou à  $4^{\circ}\text{C}$ , comprenant l'adjonction de ladite protéine à une solution comprenant :

- d'environ 1 mM à environ 10 mM de fructose 6-phosphate, notamment environ 1 mM,
- d'environ 1 mM à environ 5 mM de Tris(2-carboxyethyl)phosphine, notamment environ 1 mM,
- d'environ 5 % à environ 20 % de glycérol, notamment environ 10%.

Le fructose-6-phosphate est un substrat de ladite protéine.

Le Tris(2-carboxyethyl)phosphine est un composé réducteur permettant avantageusement de maintenir les propriétés de résines portant des ions  $\text{Ni}^{2+}$  ou  $\text{Co}^{2+}$ .

Avantageusement le glycérol est un agent cryoprotectant.

A ce titre la présente invention concerne également une composition comprenant une protéine GFAT active, le cas échéant liée à une étiquette de purification, telle

qu'une protéine telle que définie ci-dessus, ladite protéine étant susceptible de pouvoir être conservée sous une forme enzymatiquement active, durant au moins 8 jours à une température de 2°C à 10°C, notamment environ 4°C, et durant au moins 12 mois à une température de -100°C à -20°C, notamment environ -80°C, ladite protéine étant en association avec :

- environ 1 mM à environ 10 mM de fructose 6-phosphate, notamment environ 1 mM,
- environ 1 mM à environ 5 mM de Tris(2-carboxyethyl)phosphine, notamment environ 1 mM,
- environ 5 % à environ 20 % de glycérol, notamment environ 10%.

La présente invention concerne également l'utilisation d'une protéine telle que définie ci-dessus, pour le criblage de composés modifiant l'activité de ladite protéine, en particulier pour le criblage d'inhibiteur de ladite protéine.

L'activité des protéines selon l'invention peut être en particulier mesurée à l'aide des méthodes suivantes :

- la méthode radiométrique décrite par Broschat *et al.*, *Analytical Biochem.* (2002) 305:10-15,
- la méthode dite au Nitro Bleu Tétrazolium décrite par Nakata *et al.*, *J. Antibio.* (2001) 54:737-743.
- la méthode Morgan-Elson décrite par Ghosh *et al.*, *Method. Enzymol.* (1960) 5:414 et détaillée dans l'exemple qui suit.
- la méthode APAD décrite par Badet *et al.*, *Biochemistry* (1987) 26:1940 et détaillée dans l'exemple qui suit.

Avantageusement ces méthodes peuvent être utilisées pour le criblage, notamment à haut débit, de composés modifiant l'activité des protéines selon l'invention.

L'invention concerne en particulier l'utilisation telle que définie ci-dessus, pour le criblage de composés utiles dans le cadre du traitement ou de la prévention du diabète, notamment du diabète de type II, de l'obésité, de l'acidose, de la cétose, de l'arthrite, des cancers, ou de l'ostéoporose.

## DESCRIPTION DE LA FIGURE 1

La Figure 1 représente le plasmide pFastBac-gfat-His6 de poids moléculaire 6,89 kb. La cassette « Ampr » représente un gène de résistance à l'ampicilline, la cassette « ori » représente une origine de réplication bactérienne, la cassette « Gmr » représente un gène de résistance à la gentamicine, la cassette « Polh Pr » représente le promoteur de la polyhédrine, la cassette gfat-his6 représente le gène gfat1 modifié par l'insertion d'une séquence codant pour une étiquette hexahistidine. Les sites de restriction *Xba*I en position 4,11 kb, et *Eco*RI en positions 4,56 kb et 6,60 kb sont également représentés.

## EXEMPLE 1

1. Synthèse et clonage du gène *gfat1*-His6

Le fragment *EcoRI* d'un ADNc correspondant au gène *gfat1* humain a été cloné dans le site *EcoR* I du vecteur pCRII (Invitrogen) pour former le plasmide pCRII-*gfat1*. La séquence nucléotidique d'une étiquette de purification interne composée de 6 résidus histidine a été introduite en position 898 de la séquence du gène *gfat1* cloné dans pCRII par PCR avec la *Platinum pfx polymerase* (Roche) et la paire d'amorces appropriées :

- Start Aat II-His6 :

5'TGGACGTCTTTCTATCCATCGAATTAAACGAACTGCAGGACATCACC  
ATCACCATCACGATCACCCCGGACG 3' (SEQ ID NO : 14)

- End Hinc II :

5' CAAAGTTGACTCTTCCTCTCATTGTGTTCACGACAGACTCTGGC 3'  
(SEQ ID NO : 15)

selon le protocole suivant : 94°C, 2min puis 30 cycles (94°C 45 sec, 55°C 1 min, 72°C 5 min) suivis d'une polymérisation de 5 min à 72°C et retour à 4°C.

Après digestion par *AatII* et *HincII* puis purification sur gel d'agarose Seaplaque 1,5% (Tebu), l'amplicon (170 bp) a été inséré au niveau des sites de restriction correspondants dans la construction pCRII-*gfat1*. L'insert de 170 pb a été introduit par ligation dans la construction avec un rapport 3:1 à 16°C pendant la nuit en présence de T4 DNA ligase (Nebs). Le mélange de ligation (20 µl) ainsi obtenu a permis de transformer une souche d'*E. coli* JM109. Ensuite, le fragment *XbaI*-*HindIII* du plasmide recombinant pCRII-*gfat1*-His6 a été cloné dans le plasmide donneur pFastBac1 (Life Technologies Ltd). Le plasmide pFastBac-*gfat*-His6 ainsi généré (**Figure 1**) a été vérifié par digestions multiples: *SmaI*, *AccI/DraI*, *PstEI/XbaI*, et par séquençage. En vue d'améliorer la construction, la séquence en amont du codon de démarrage a été mutée sur deux positions par PCR, avec la paire d'amorces suivante, afin d'enlever deux phases de lecture ouvertes en amont du gène *gfat1* :

- Start *XbaI*

5' AATCTAGATTCATGCTCGAGCGGCCGCCAGTGTGATTGATATC 3'  
(SEQ ID NO : 16)

- End AfeI

5' ATTTTATCAGAGCGCTGGGGGTGGCTATTGACAGG 3' (SEQ ID NO : 17)

selon le protocole : 94°C 2 min, puis 30 cycles (94°C 15 sec, 55°C 30 sec, 68°C 1 min) suivis d'une polymérisation de 1 min à 68°C et retour à 4°C.

Le fragment de PCR obtenu, contenant les deux mutations, a été purifié sur gel SeaPlaque (Tebu) à 0,7% puis digéré par *XbaI* et *AfeI* pour remplacer son homologue dans pFastBac-gfat-His6 afin de donner le plasmide donneur pFastBac-gfat-His6-2orf servant à la transposition dans les cellules DH10Bac (Life Technologies Ltd). La construction a été vérifiée par digestions *SmaI*, *XbaI/PstEI*, *XbaI/HindIII*, et par séquençage.

Un bacmide recombinant a été isolé après transposition dans les cellules DH10Bac et utilisé pour transfecter des cellules d'insecte Sf9 en présence de Lipofectin (Life Technologies Ltd). Les baculovirus obtenus ont été amplifiés dans les cellules Sf9 et le titre viral a été mesuré à  $5.10^7$  pfu/ml.

### 2. Production de la protéine GFAT1-His6

Des cellules d'insectes Sf9 ont été cultivées à 28°C en présence de milieu SF900II (Life Technologies Ltd) en fioles de 5 L sous agitation à 100 rpm. Les cellules à une densité de  $2.10^9$  cellules /L ont été infectées par le baculovirus recombinant obtenu ci-dessus avec une multiplicité d'infection de 2 (pfu/cellule), puis cultivées durant 72 h.

Les cellules et le surnageant ont été séparés par centrifugation (2500 g, 10 min à 4°C). Les culots cellulaires ont été lavés en présence de tampon Tris-HCl 20 mM, pH 7, centrifugés (4000 g, 45 min à 4°C) et congelés à -80°C.

### 3. Purification de la protéine GFAT1-His6

Le culot cellulaire (20 g) a été repris dans 50 ml de tampon de lyse (NaPO<sub>4</sub> 50mM pH 7,5, NaCl 300 mM, imidazole 10 mM, fructose-6-phosphate (fructose-6P) 1 mM, TCEP (Tris(2-carboxyethyl)phosphine) 1 mM, PMSF 1 mM (fluorure de phénylmethylsulfonyl), 10% glycérol et 1 tablette de cocktail d'inhibiteurs de protéases sans EDTA (Roche Applied Sciences) et soumis à un broyage au DynoMill à 4500 trs/min (4 cycles de 30 sec) en présence de 40 g de microbilles de 0,2 mm de diamètre. L'ensemble a été refroidi par circulation d'éthylène glycol/eau réglé à -15°C. L'extrait brut obtenu (100 ml, 445 mg de protéines totales) a été centrifugé à 4°C pendant 20 min

à 12000 tpm. Le surnageant a été soumis à une ultracentrifugation à 4°C (350000 tpm, 1h). Le surnageant ainsi obtenu a été mélangé à 5 ml de matrice 50% Ni-NTA (Qiagen) pendant 2 h à 4°C. Le mélange a été coulé dans une colonne vide puis rincé avec 40 ml de tampon de lavage (NaPO<sub>4</sub> 50 mM pH 7,5, NaCl 300 mM, imidazole 40 mM, fructose-6P 1 mM, TCEP 1 mM, PMSF 1 mM et 1 tablette de cocktail d'inhibiteurs de protéases sans EDTA (Roche Applied Sciences). L'élution a été réalisée par paliers successifs à 125 et 500 mM imidazole dans le même tampon que précédemment. 12 mg de GFAT1-His6 fonctionnelle (dosage protéique selon la méthode de Bradford ont ainsi été obtenus.

#### 4. Conservation de l'enzyme GFAT1-His6

L'enzyme a alors été stockée en fractions de 100 µl en présence de 1 mM fructose-6P, 1mM TCEP et 10% glycérol à -80°C. La stabilité de l'enzyme est de plusieurs mois à -80°C et supérieure à 8 jours à 4°C.

#### 5. Dosage de l'activité de l'enzyme GFAT1-His6

Différents tests de dosage de l'activité enzymatique de la GFAT1-His6 ont été utilisés. Ces tests peuvent être également utilisés afin de cribler des composés modifiant, et notamment inhibant, l'activité de la GFAT1-His6. Il est possible des les adapter aisément à un criblage à haut débit.

##### *Dosage de Morgan-Elson :*

Dans ce cas l'activité enzymatique est suivie par un test colorimétrique dont le principe est le suivant : la D-glucosamine-6P libérée par l'enzyme est N-acétylée par l'anhydride acétique en milieu alcalin (Ghosh *et al.*, *Method. Enzymol.* (1962) 5:414), puis La solution est traitée par le réactif d'Ehrlich (para-diméthyl-amino-benzaldéhyde, PDAB) en milieu acide concentré ; le composé rose formé absorbe à 585 nm.

La réaction enzymatique se déroule pendant 30 min à 37°C en présence de :

- 0,2 ml de fructose-6P à 100 mM
- 0,25 ml de L-Glutamine à 60 mM
- 0,25 ml de tampon KPO<sub>4</sub> à 150 mM pH 7
- 0,1 ml d'EDTA (éthylène diamine tétra-acétate) à 25 mM, pH 7
- jusqu'à 200 µl d'échantillon (à compléter avec H<sub>2</sub>O si nécessaire)



La réaction est stoppée par immersion 4 minutes dans un bain d'eau à 100°C puis centrifugée. 0,8 ml du surnageant sont prélevés pour le dosage de la glucosamine-6P suivant le protocole suivant :

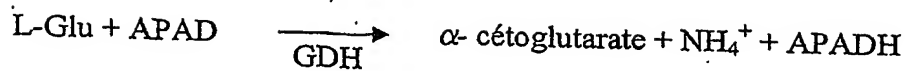
- addition de 0,1 ml de NaHCO<sub>3</sub> saturé,
- addition de 0,1 ml d'une solution d'anhydride acétique à 5 % dans l'eau préparée extemporanément,
- agitation et incubation 5 min à température ambiante,
- incubation 5 min dans un bain à 100°C,
- addition de 0,2 ml de borate de potassium 0,8 M pH 9,1 (à ajuster avec KOH 10 N).
- agitation et incubation 7 min dans un bain à 100°C.
- addition de 3 ml de réactif d'Ehrlich dilué 10 fois dans l'acide acétique, préparé extemporanément, sur la solution refroidie dans la glace,
- incubation 20 min à 37°C.

L'activité de la GFAT a été déterminée comparativement à une courbe étalon établie en utilisant la D-glucosamine comme standard dans une gamme de concentration de 0 à 200 nmoles. L'activité spécifique de la GFAT1-His6 obtenue a été ainsi mesurée à 1,7 U/mg. Ce qui est supérieur à la valeur de 0,4 U/mg obtenue par Broschat *et al.*, *J. Biol. Chem.* (2002) 277:14764, pour la purification d'une GFAT1 humaine recombinante. Ceci traduit une activité supérieure de la GFAT1-His6 et/ou une pureté plus importante de la préparation enzymatique selon l'invention.

Les paramètres cinétiques de la GFAT1-His6 ont été caractérisés vis-à-vis de la glutamine ( $K_m^{\text{Gln}} = 0,2 \text{ mM}$ ) et du fructose-6P (F6P) ( $K_m^{\text{F6P}} = 0,006 \text{ mM}$ ) par un dosage spectrophotométrique couplé à la glutamate déshydrogénase selon le test APAD. Ce qui est conforme aux valeurs citées dans l'art antérieur ( $K_m^{\text{Gln}} = 0,26 \text{ mM}$  et  $K_m^{\text{F6P}} = 0,007 \text{ mM}$  pour Broschat *et al.*, *J. Biol. Chem.* (2002) 277:14764).

#### *Dosage APAD*

Il s'agit d'un dosage spectrophotométrique dans l'ultraviolet de l'activité GFAT. Il est basé sur la détermination, en continu, de la quantité de L-glutamate formé à l'aide de la GFAT et d'un analogue du NAD (nicotinamide adénine dinucléotide), l'APAD (3-acetylpyridine adenine dinucleotide), selon la réaction suivante (catalysée par la glutamate déshydrogénase (GDH)) :



La mesure s'effectue à 365 nm, à 37°C. Dans ces conditions une unité d'absorbance correspond à 0,11 µmole d'APADH formé.

L'essai comprend :

- 100 µl APAD 3 mM (2 mg/ml)
- 25 µl KCl 2M
- 100 µl de tampon KPO<sub>4</sub> 1 M pH 7,2
- 100 µl de Fructose-6P 100 mM (30,41 mg/ml)
- 100 µl de L-Glutamine 60 mM purifiée (8,77 mg/ml)
- H<sub>2</sub>O qsp 1 ml (en tenant compte des volumes à rajouter ci-après)
- 50 µl GDH
- échantillon à doser : 0,5 µg

Il est également possible d'utiliser d'autres procédés de dosage, tels que le dosage radiométrique décrit par Broschat *et al.*, *Analytical Biochem.* (2002) 305:10-15 ou le dosage dit Nitro Bleu Tétrazolium décrit par Nakata *et al.*, *J. Antibio.* (2001) 54:737-743.

## REVENDICATIONS

1. Protéine enzymatiquement active comprenant :

- une séquence de GFAT et au moins une séquence d'une étiquette de purification, la séquence de l'étiquette de purification étant insérée entre deux acides aminés consécutifs de la séquence de GFAT, ou
- une séquence dérivant de la séquence précédente par suppression, insertion ou mutation d'au moins un acide aminé, sous réserve que ladite protéine présente une activité enzymatique, ou
- une séquence présentant au moins 35 % d'identité de séquence et/ou au moins 44 % de similarité de séquence avec l'une des séquences précédentes, sous réserve que ladite protéine présente une activité enzymatique.

2. Protéine selon la revendication 1, dans laquelle la séquence de GFAT correspond à une séquence de GFAT de bactérie ou d'eucaryote, notamment de plante, de champignon ou d'animal, en particulier d'insecte ou de mammifère, plus particulièrement d'homme.

3. Protéine selon la revendication 1 ou 2, dans laquelle la séquence de l'étiquette de purification est insérée entre deux acides aminés consécutifs de la séquence de GFAT, lesdits acides aminés étant compris dans :

- une partie de la séquence de la GFAT correspondant à la séquence s'étendant entre le feuillet  $\beta 2$  et le feuillet  $\beta 3$  de la GFAT d'*Escherichia coli*, ou
- une partie de la séquence de la GFAT correspondant à la séquence s'étendant entre le feuillet  $\beta 13$  et le feuillet  $\beta 14$  de la GFAT d'*Escherichia coli*, ou
- une partie de la séquence de la GFAT correspondant à la séquence s'étendant entre le feuillet  $\beta 15$  et l'hélice  $\alpha 6$  de la GFAT d'*Escherichia coli*.

4. Protéine selon l'une des revendications 1 à 3, dans laquelle la séquence de GFAT correspond à :

- SEQ ID NO : 2, correspondant à la séquence de la GFAT1 humaine,
- SEQ ID NO : 4, correspondant à la séquence de la GFAT2 humaine,
- SEQ ID NO : 6, correspondant à la séquence de la GFAT1Alt humaine.

## REVENDICATIONS

1. Protéine correspondant à une glutamine:fructose-6-phosphate amidotransférase (GFAT) modifiée enzymatiquement active d'origine humaine, ladite GFAT étant modifiée de telle sorte qu'elle comprend au moins une séquence d'une étiquette de purification correspondant à une séquence peptidique d'environ 2 à environ 10 acides aminés, notamment d'environ 4 à environ 8 acides aminés, susceptible de se lier spécifiquement à un ligand donné, ce qui permet la formation d'un complexe entre la GFAT portant l'étiquette de purification et ledit ligand, ledit complexe pouvant être spécifiquement isolé, la séquence de l'étiquette de purification étant insérée entre deux acides aminés consécutifs de la séquence de GFAT, ces deux acides aminés étant compris entre les acides aminés :

- 43 à 47, 298 à 306, et/ou 342 à 347 de SEQ ID NO : 2 de la GFAT1 humaine,
- 42 à 45, 299 à 307, et/ou 343 à 348 de SEQ ID NO : 4 de la GFAT2 humaine,
- 43 à 47, 316 à 324, et/ou 360 à 365 de SEQ ID NO : 6 de la GFAT1Alt humaine.

2. Protéine selon la revendication 1, dans laquelle la séquence de l'étiquette de purification est insérée entre les acides aminés :

- 299 et 300 de SEQ ID NO : 2.
- 300 et 301 de SEQ ID NO : 4
- 317 et 318 de SEQ ID NO : 6

3. Protéine selon la revendication 1 ou 2, dans laquelle l'étiquette de purification est une hexa-histidine.

4. Protéine selon l'une des revendications 1 à 3 correspondant aux séquences :

- SEQ ID NO : 8, correspondant à la séquence SEQ ID NO : 2 dans laquelle une hexa-histidine est insérée entre les acides aminés 299 et 300,
- SEQ ID NO : 10, correspondant à la séquence SEQ ID NO : 4 dans laquelle une hexa-histidine est insérée entre les acides aminés 300 et 301, et
- SEQ ID NO : 12, correspondant à la séquence SEQ ID NO : 6 dans laquelle une hexa-histidine est insérée entre les acides aminés 317 et 318.

5. Protéine selon l'une des revendications 1 à 4, dans laquelle la séquence de l'étiquette de purification est insérée entre deux acides aminés consécutifs, lesdits acides aminés étant compris entre les acides aminés :

- 43 à 47, 298 à 306, et/ou 342 à 347 de SEQ ID NO : 2

- 42 à 45, 299 à 307, et/ou 343 à 348 de SEQ ID NO : 4

- 43 à 47, 316 à 324, et/ou 360 à 365 de SEQ ID NO : 6

6. Protéine selon la revendication 5, dans laquelle la séquence de l'étiquette de purification est insérée entre les acides aminés :

- 299 et 300 de SEQ ID NO : 2.

- 300 et 301 de SEQ ID NO : 4

- 317 et 318 de SEQ ID NO : 6

7. Protéine selon l'une des revendications 1 à 6, dans laquelle l'étiquette de purification correspond à une séquence d'environ 2 à environ 10 acides aminés, notamment d'environ 4 à environ 8 acides aminés.

8. Protéine selon l'une des revendications 1 à 7, dans laquelle l'étiquette de purification est une hexa-histidine.

9. Protéine selon les revendications 6 et 8 correspondant aux séquences :

- SEQ ID NO : 8, correspondant à la séquence SEQ ID NO : 2 dans laquelle une hexa-histidine est insérée entre les acides aminés 299 et 300,

- SEQ ID NO : 10, correspondant à la séquence SEQ ID NO : 4 dans laquelle une hexa-histidine est insérée entre les acides aminés 300 et 301, et

- SEQ ID NO : 12, correspondant à la séquence SEQ ID NO : 6 dans laquelle une hexa-histidine est insérée entre les acides aminés 317 et 318.

10. Acide nucléique comprenant ou étant constitué d'une séquence codant pour une protéine selon l'une des revendications 1 à 9.

11. Acide nucléique comprenant ou étant constitué de la séquence nucléotidique :

- SEQ ID NO : 7 codant pour la protéine SEQ ID NO : 8, ou

- SEQ ID NO : 9 codant pour la protéine SEQ ID NO : 10, ou

5. Acide nucléique comprenant ou étant constitué d'une séquence codant pour une protéine selon l'une des revendications 1 à 3.

6. Acide nucléique comprenant ou étant constitué de la séquence nucléotidique :

- SEQ ID NO : 7 codant pour la protéine SEQ ID NO : 8, ou
- SEQ ID NO : 9 codant pour la protéine SEQ ID NO : 10, ou
- SEQ ID NO : 11 codant pour la protéine SEQ ID NO : 12,

ou de son complémentaire, ou étant dérivé de ladite séquence par mutation, insertion ou délétion d'au moins un nucléotide, sous réserve que ladite séquence nucléotidique code pour une protéine enzymatiquement active.

7. Vecteur eucaryote ou procaryote comprenant un acide nucléique selon la revendication 5 ou 6.

8. Procédé de purification d'une protéine selon l'une des revendications 1 à 3, à partir d'une solution comprenant ladite protéine, comprenant une étape de mise en présence de ladite solution avec un composé se liant spécifiquement à l'étiquette de purification de ladite protéine et une étape de séparation du complexe formé par la liaison de ladite protéine audit composé des autres constituants de la solution.

9. Procédé de purification selon la revendication 10, comprenant une étape de mise en présence d'une solution comprenant une protéine selon la revendication 3 avec un composé comprenant un cation métallique divalent tel que  $\text{Ni}^{2+}$  ou  $\text{Co}^{2+}$ , notamment  $\text{Ni}^{2+}$ , et une étape de séparation du complexe formé par la liaison de la protéine audit composé des autres constituants de la solution.

10. Procédé de conservation d'une protéine selon l'une des revendications 1 à 3 sous une forme enzymatiquement active, notamment à  $-80^{\circ}\text{C}$  ou à  $4^{\circ}\text{C}$ , comprenant l'adjonction de ladite protéine à une solution comprenant :

- d'environ 1 mM à environ 10 mM de fructose 6-phosphate, notamment environ 1 mM,
- d'environ 1 mM à environ 5 mM de Tris(2-carboxyethyl)phosphine, notamment environ 1 mM,
- d'environ 5 % à environ 20 % de glycérol, notamment environ 10%.

- SEQ ID NO : 11 codant pour la protéine SEQ ID NO : 12,  
ou de son complémentaire, ou étant dérivé de ladite séquence par mutation, insertion ou  
délétion d'au moins un nucléotide, sous réserve que ladite séquence nucléotidique code  
pour une protéine enzymatiquement active.

12. Vecteur eucaryote ou procaryote comprenant un acide nucléique selon la  
revendication 10 ou 11.

13. Procédé de purification d'une protéine selon l'une des revendications 1 à 9, à partir  
d'une solution comprenant ladite protéine, comprenant une étape de mise en présence  
de ladite solution avec un composé se liant spécifiquement à l'étiquette de purification  
de ladite protéine et une étape de séparation du complexe formé par la liaison de ladite  
protéine audit composé des autres constituants de la solution.

14. Procédé de purification selon la revendication 13, comprenant une étape de mise en  
présence d'une solution comprenant une protéine selon la revendication 9 avec un  
composé comprenant un cation métallique divalent tel que  $\text{Ni}^{2+}$  ou  $\text{Co}^{2+}$ , notamment  
 $\text{Ni}^{2+}$ , et une étape de séparation du complexe formé par la liaison de la protéine audit  
composé des autres constituants de la solution.

15. Procédé de conservation d'une protéine selon l'une des revendications 1 à 9 sous  
une forme enzymatiquement active, notamment à  $-80^{\circ}\text{C}$  ou à  $4^{\circ}\text{C}$ , comprenant  
l'adjonction de ladite protéine à une solution comprenant :

- d'environ 1 mM à environ 10 mM de fructose 6-phosphate, notamment environ  
1 mM,
- d'environ 1 mM à environ 5 mM de Tris(2-carboxyethyl)phosphine,  
notamment environ 1 mM,
- d'environ 5 % à environ 20 % de glycérol, notamment environ 10%.

16. Composition comprenant une protéine GFAT active, le cas échéant liée à une  
étiquette de purification, telle qu'une protéine selon l'une des revendication 1 à 9, ladite  
protéine étant susceptible de pouvoir être conservée sous une forme enzymatiquement  
active, durant au moins 8 jours à une température de  $2^{\circ}\text{C}$  à  $10^{\circ}\text{C}$ , notamment environ

11. Composition comprenant une protéine GFAT active liée à une étiquette de purification selon l'une des revendication 1 à 3, ladite protéine étant susceptible de pouvoir être conservée sous une forme enzymatiquement active, durant au moins 8 jours à une température de 2°C à 10°C, notamment environ 4°C, et durant au moins 12 mois à une température de -100°C à -20°C, notamment environ -80°C, ladite protéine étant en association avec :

- environ 1 mM à environ 10 mM de fructose 6-phosphate, notamment environ 1 mM,
- environ 1 mM à environ 5 mM de Tris(2-carboxyethyl)phosphine, notamment environ 1 mM,
- environ 5 % à environ 20 % de glycérol, notamment environ 10%.

12. Utilisation d'une protéine selon l'une des revendications 1 à 3, pour le criblage de composés modifiant l'activité de ladite protéine, en particulier pour le criblage d'inhibiteur de ladite protéine.

13. Utilisation selon la revendication 12, pour le criblage de composés utiles dans le cadre du traitement ou de la prévention du diabète, notamment du diabète de type II, de l'obésité, de l'acidose, de la cétose, de l'arthrite, des cancers, ou de l'ostéoporose.



4°C, et durant au moins 12 mois à une température de -100°C à -20°C, notamment environ -80°C, ladite protéine étant en association avec :

- environ 1 mM à environ 10 mM de fructose 6-phosphate, notamment environ 1 mM,
- environ 1 mM à environ 5 mM de Tris(2-carboxyethyl)phosphine, notamment environ 1 mM,
- environ 5 % à environ 20 % de glycérol, notamment environ 10%.

5

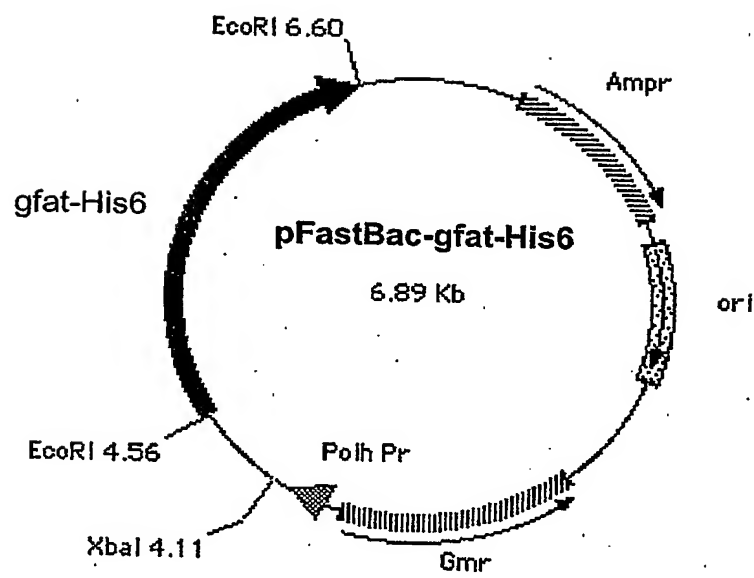
10

17. Utilisation d'une protéine selon l'une des revendications 1 à 9, pour le criblage de composés modifiant l'activité de ladite protéine, en particulier pour le criblage d'inhibiteur de ladite protéine.

15

18. Utilisation selon la revendication 17, pour le criblage de composés utiles dans le cadre du traitement ou de la prévention du diabète, notamment du diabète de type II, de l'obésité, de l'acidose, de la cétose, de l'arthrite, des cancers, ou de l'ostéoporose.

5



10

15

**Figure 1**

# LISTE DE SEQUENCES

<110> CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

<120> GLUTAMINE:FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE (GFAT) COMPRENANT UNE ETIQUETTE DE PURIFICATION INTERNE, ET SON UTILISATION POUR LE CRIBLAGE DE COMPOSES

<130> IFB 03 BP CNR GFAT

<160> 19

<170> PatentIn version 3.1

<210> 1

<211> 2046

<212> ADN

<213> Homo sapiens

<220>

<221> CDS

<222> (1)..(2046)

<223>

<220>

<221> misc\_feature

<222> (170)..(170)

<223> t ou c

<400> 1

atg	tgt	ggt	ata	ttt	gct	tac	tta	aac	tac	cat	gtt	cct	cga	acg	aga		
Met	Cys	Gly	Ile	Phe	Ala	Tyr	Leu	Asn	Tyr	His	Val	Pro	Arg	Thr	Arg		48
1			5						10					15			

cga	gaa	atc	ctg	gag	acc	cta	atc	aaa	ggc	ctt	cag	aga	ctg	gag	tac		
Arg	Glu	Ile	Leu	Glu	Thr	Leu	Ile	Lys	Gly	Leu	Gln	Arg	Leu	Glu	Tyr		96
		20						25					30				

aga	gga	tat	gat	tct	gct	ggt	gtg	gga	ttt	gat	gga	ggc	aat	gat	aaa		
Arg	Gly	Tyr	Asp	Ser	Ala	Gly	Val	Gly	Phe	Asp	Gly	Gly	Asn	Asp	Lys		144
		35				40						45					

gat	tgg	gaa	gcc	aat	gcc	tgc	aaa	anc	cag	ctt	att	aag	aag	aaa	gga		
Asp	Trp	Glu	Ala	Asn	Ala	Cys	Lys	Xaa	Gln	Leu	Ile	Lys	Lys	Lys	Gly		192
	50				55						60						

aaa	ggt	aag	gca	ctg	gat	gaa	gaa	gtt	cac	aag	caa	caa	gat	atg	gat		
Lys	Val	Lys	Ala	Leu	Asp	Glu	Glu	Val	His	Lys	Gln	Gln	Asp	Met	Asp		240
65				70					75					80			

ttg	gat	ata	gaa	ttt	gat	gta	cac	ctt	gga	ata	gct	cat	acc	cgt	tgg		
Leu	Asp	Ile	Glu	Phe	Asp	Val	His	Leu	Gly	Ile	Ala	His	Thr	Arg	Trp		288
			85					90						95			

gca	aca	cat	gga	gaa	ccc	agt	cct	gtc	aat	agc	cac	ccc	cag	cgc	tct		
Ala	Thr	His	Gly	Glu	Pro	Ser	Pro	Val	Asn	Ser	His	Pro	Gln	Arg	Ser		336
			100					105					110				

gat	aaa	aat	aat	gaa	ttt	atc	gtt	att	cac	aat	gga	atc	atc	acc	aac		
Asp	Lys	Asn	Asn	Glu	Phe	Ile	Val	Ile	His	Asn	Gly	Ile	Ile	Thr	Asn		384
		115				120						125					

tac	aaa	gac	ttg	aaa	aag	ttt	ttg	gaa	agc	aaa	ggc	tat	gac	ttc	gaa		
Tyr	Lys	Asp	Leu	Lys	Lys	Phe	Leu	Glu	Ser	Lys	Gly	Tyr	Asp	Phe	Glu		432
	130					135						140					

tct Ser 145	gaa Glu	aca Thr	gac Asp	aca Thr	gag Glu	aca Thr	att Ile	gcc Ala	aag Lys	ctc Leu	gtt Val	aag Lys	tat Tyr	atg Met	tat Tyr	480
					150					155					160	
gac Asp	aat Asn	cgg Arg	gaa Glu	agt Ser	caa Gln	gat Asp	acc Thr	agc Ser	ttt Phe	act Thr	acc Thr	ttg Leu	gtg Val	gag Glu	aga Arg	528
				165					170					175		
gtt Val	atc Ile	caa Gln	caa Gln	ttg Leu	gaa Glu	ggg Gly	gct Ala	ttt Phe	gca Ala	ctt Leu	gtg Val	ttt Phe	aaa Lys	agt Ser	gtt Val	576
			180					185					190			
cat His	ttt Phe	ccc Pro	ggg Gly	caa Gln	gca Ala	gtt Val	ggc Gly	aca Thr	agg Arg	cga Arg	ggg Gly	agc Ser	cct Pro	ctg Leu	ttg Leu	624
		195					200					205				
att Ile	ggg Gly	gta Val	cgg Arg	agt Ser	gaa Glu	cat His	aaa Lys	ctt Leu	tct Ser	act Thr	gat Asp	cac His	att Ile	cct Pro	ata Ile	672
	210					215					220					
ctc Leu	tac Tyr	aga Arg	aca Thr	ggc Gly	aaa Lys	gac Asp	aag Lys	aaa Lys	gga Gly	agc Ser	tgc Cys	aat Asn	ctc Leu	tct Ser	cgt Arg	720
	225				230					235					240	
gtg Val	gac Asp	agc Ser	aca Thr	acc Thr	tgc Cys	ctt Leu	ttc Phe	ccg Pro	gtg Val	gaa Glu	gaa Glu	aaa Lys	gca Ala	gtg Val	gag Glu	768
				245				250						255		
tat Tyr	tac Tyr	ttt Phe	gct Ala	tct Ser	gat Asp	gca Ala	agt Ser	gct Ala	gtc Val	ata Ile	gaa Glu	cac His	acc Thr	aat Asn	cgc Arg	816
			260					265					270			
gtc Val	atc Ile	ttt Phe	ctg Leu	gaa Glu	gat Asp	gat Asp	gat Asp	gtt Val	gca Ala	gca Ala	gta Val	gtg Val	gat Asp	gga Gly	cgt Arg	864
		275					280					285				
ctt Leu	tct Ser	atc Ile	cat His	cga Arg	att Ile	aaa Lys	cga Arg	act Thr	gca Ala	gga Gly	gat Asp	cac His	ccc Pro	gga Gly	cga Arg	912
	290					295					300					
gct Ala	gtg Val	caa Gln	aca Thr	ctc Leu	cag Gln	atg Met	gaa Glu	ctc Leu	cag Gln	cag Gln	atc Ile	atg Met	aag Lys	ggc Gly	aac Asn	960
	305				310				315						320	
ttc Phe	agt Ser	tca Ser	ttt Phe	atg Met	cag Gln	aag Lys	gaa Glu	ata Ile	ttt Phe	gag Glu	cag Gln	cca Pro	gag Glu	tct Ser	gtc Val	1008
				325					330					335		
gtg Val	aac Asn	aca Thr	atg Met	aga Arg	gga Gly	aga Arg	gtc Val	aac Asn	ttt Phe	gat Asp	gac Asp	tat Tyr	act Thr	gtg Val	aat Asn	1056
			340					345					350			
ttg Leu	ggg Gly	ggg Gly	ttg Leu	aag Lys	gat Asp	cac His	ata Ile	aag Lys	gag Glu	atc Ile	cag Gln	aga Arg	tgc Cys	cgg Arg	cgt Arg	1104
	355						360					365				
ttg Leu	att Ile	ctt Leu	att Ile	gct Ala	tgt Cys	gga Gly	aca Thr	agt Ser	tac Tyr	cat His	gct Ala	ggg Gly	gta Val	gca Ala	aca Thr	1152
	370					375					380					
cgt Arg	caa Gln	gtt Val	ctt Leu	gag Glu	gag Glu	ctg Leu	act Thr	gag Glu	ttg Leu	cct Pro	gtg Val	atg Met	gtg Val	gaa Glu	cta Leu	1200
	385				390				395					400		
gca Val	agt Ser	gac Ser	ttc Ser	ctg Ser	gac Ser	aga Ser	aac Ser	aca Ser	cca Ser	gtc Ser	ttt Ser	cga Ser	gat Ser	gat Ser	gtt Ser	1248

Ala	Ser	Asp	Phe	Leu	Asp	Arg	Asn	Thr	Pro	Val	Phe	Arg	Asp	Asp	Val	
				405					410					415		
tgc	ttt	ttc	ctt	agt	caa	tca	ggt	gag	aca	gca	gat	act	ttg	atg	ggt	1296
Cys	Phe	Phe	Leu	Ser	Gln	Ser	Gly	Glu	Thr	Ala	Asp	Thr	Leu	Met	Gly	
			420					425					430			
ctt	cgt	tac	tgt	aag	gag	aga	gga	gct	tta	act	gtg	ggg	atc	aca	aac	1344
Leu	Arg	Tyr	Cys	Lys	Glu	Arg	Gly	Ala	Leu	Thr	Val	Gly	Ile	Thr	Asn	
		435					440					445				
aca	gtt	ggc	agt	tcc	ata	tca	cgg	gag	aca	gat	tgt	gga	gtt	cat	att	1392
Thr	Val	Gly	Ser	Ser	Ile	Ser	Arg	Glu	Thr	Asp	Cys	Gly	Val	His	Ile	
	450					455					460					
aat	gct	ggc	cct	gag	att	ggt	gtg	gcc	agt	aca	aag	gct	tat	acc	agc	1440
Asn	Ala	Gly	Pro	Glu	Ile	Gly	Val	Ala	Ser	Thr	Lys	Ala	Tyr	Thr	Ser	
465					470				475					480		
cag	ttt	gta	tcc	ctt	gtg	atg	ttt	gcc	ctt	atg	atg	tgt	gat	gat	cgg	1488
Gln	Phe	Val	Ser	Leu	Val	Met	Phe	Ala	Leu	Met	Met	Cys	Asp	Asp	Arg	
				485				490						495		
atc	tcc	atg	caa	gaa	aga	cgc	aaa	gag	atc	atg	ctt	gga	ttg	aaa	cgg	1536
Ile	Ser	Met	Gln	Glu	Arg	Arg	Lys	Glu	Ile	Met	Leu	Gly	Leu	Lys	Arg	
			500					505					510			
ctg	cct	gat	ttg	att	aag	gaa	gta	ctg	agc	atg	gat	gac	gaa	att	cag	1584
Leu	Pro	Asp	Leu	Ile	Lys	Glu	Val	Leu	Ser	Met	Asp	Asp	Glu	Ile	Gln	
		515					520					525				
aaa	cta	gca	aca	gaa	ctt	tat	cat	cag	aag	tca	gtt	ctg	ata	atg	gga	1632
Lys	Leu	Ala	Thr	Glu	Leu	Tyr	His	Gln	Lys	Ser	Val	Leu	Ile	Met	Gly	
	530					535					540					
cga	ggc	tat	cat	tat	gct	act	tgt	ctt	gaa	ggg	gca	ctg	aaa	atc	aaa	1680
Arg	Gly	Tyr	His	Tyr	Ala	Thr	Cys	Leu	Glu	Gly	Ala	Leu	Lys	Ile	Lys	
545					550					555				560		
gaa	att	act	tat	atg	cac	tct	gaa	ggc	atc	ctt	gct	ggt	gaa	ttg	aaa	1728
Glu	Ile	Thr	Tyr	Met	His	Ser	Glu	Gly	Ile	Leu	Ala	Gly	Glu	Leu	Lys	
				565					570					575		
cat	ggc	cct	ctg	gct	ttg	gtg	gat	aaa	ttg	atg	cct	gtg	atc	atg	atc	1776
His	Gly	Pro	Leu	Ala	Leu	Val	Asp	Lys	Leu	Met	Pro	Val	Ile	Met	Ile	
			580					585					590			
atc	atg	aga	gat	cac	act	tat	gcc	aag	tgt	cag	aat	gct	ctt	cag	caa	1824
Ile	Met	Arg	Asp	His	Thr	Tyr	Ala	Lys	Cys	Gln	Asn	Ala	Leu	Gln	Gln	
		595					600					605				
gtg	gtt	gct	cgg	cag	ggg	cgg	cct	gtg	gta	att	tgt	gat	aag	gag	gat	1872
Val	Val	Ala	Arg	Gln	Gly	Arg	Pro	Val	Val	Ile	Cys	Asp	Lys	Glu	Asp	
	610					615					620					
act	gag	acc	att	aag	aac	aca	aaa	aga	acg	atc	aag	gtg	ccc	cac	tca	1920
Thr	Glu	Thr	Ile	Lys	Asn	Thr	Lys	Arg	Thr	Ile	Lys	Val	Pro	His	Ser	
625					630					635				640		
gtg	gac	tgc	ttg	cag	ggc	att	ctc	agc	gtg	atc	cct	tta	cag	ttg	ctg	1968
Val	Asp	Cys	Leu	Gln	Gly	Ile	Leu	Ser	Val	Ile	Pro	Leu	Gln	Leu	Leu	
				645					650					655		
gct	ttc	cac	ctt	gct	gtg	ctg	aga	ggc	tat	gat	gtt	gat	ttc	cca	cgg	2016
Ala	Phe	His	Leu	Ala	Val	Leu	Arg	Gly	Tyr	Asp	Val	Asp	Phe	Pro	Arg	
			660					665					670			

aat ctt gcc aaa tct gtg act gta gag tga  
Asn Leu Ala Lys Ser Val Thr Val Glu  
675 680

<210> 2  
<211> 681  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<222> (57)..(57)  
<223> Le 'Xaa' en position 57 représente Thr ou Ile.

<400> 2

Met Cys Gly Ile Phe Ala Tyr Leu Asn Tyr His Val Pro Arg Thr Arg  
1 5 10 15

Arg Glu Ile Leu Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr  
20 25 30

Arg Gly Tyr Asp Ser Ala Gly Val Gly Phe Asp Gly Gly Asn Asp Lys  
35 40 45

Asp Trp Glu Ala Asn Ala Cys Lys Xaa Gln Leu Ile Lys Lys Lys Gly  
50 55 60

Lys Val Lys Ala Leu Asp Glu Glu Val His Lys Gln Gln Asp Met Asp  
65 70 75 80

Leu Asp Ile Glu Phe Asp Val His Leu Gly Ile Ala His Thr Arg Trp  
85 90 95

Ala Thr His Gly Glu Pro Ser Pro Val Asn Ser His Pro Gln Arg Ser  
100 105 110

Asp Lys Asn Asn Glu Phe Ile Val Ile His Asn Gly Ile Ile Thr Asn  
115 120 125

Tyr Lys Asp Leu Lys Lys Phe Leu Glu Ser Lys Gly Tyr Asp Phe Glu  
130 135 140

Ser Glu Thr Asp Thr Glu Thr Ile Ala Lys Leu Val Lys Tyr Met Tyr  
145 150 155 160

Asp Asn Arg Glu Ser Gln Asp Thr Ser Phe Thr Thr Leu Val Glu Arg  
165 170 175

Val Ile Gln Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val  
180 185 190

His Phe Pro Gly Gln Ala Val Gly Thr Arg Arg Gly Ser Pro Leu Leu  
195 200 205

Ile Gly Val Arg Ser Glu His Lys Leu Ser Thr Asp His Ile Pro Ile  
210 215 220

Leu Tyr Arg Thr Gly Lys Asp Lys Lys Gly Ser Cys Asn Leu Ser Arg  
225 230 235 240

Val Asp Ser Thr Thr Cys Leu Phe Pro Val Glu Glu Lys Ala Val Glu  
245 250 255

Tyr Tyr Phe Ala Ser Asp Ala Ser Ala Val Ile Glu His Thr Asn Arg  
260 265 270

Val Ile Phe Leu Glu Asp Asp Asp Val Ala Ala Val Val Asp Gly Arg  
275 280 285

Leu Ser Ile His Arg Ile Lys Arg Thr Ala Gly Asp His Pro Gly Arg  
290 295 300

Ala Val Gln Thr Leu Gln Met Glu Leu Gln Gln Ile Met Lys Gly Asn  
305 310 315 320

Phe Ser Ser Phe Met Gln Lys Glu Ile Phe Glu Gln Pro Glu Ser Val  
325 330 335

Val Asn Thr Met Arg Gly Arg Val Asn Phe Asp Asp Tyr Thr Val Asn  
340 345 350

Leu Gly Gly Leu Lys Asp His Ile Lys Glu Ile Gln Arg Cys Arg Arg  
355 360 365

Leu Ile Leu Ile Ala Cys Gly Thr Ser Tyr His Ala Gly Val Ala Thr  
370 375 380

Arg Gln Val Leu Glu Glu Leu Thr Glu Leu Pro Val Met Val Glu Leu  
385 390 395 400

Ala Ser Asp Phe Leu Asp Arg Asn Thr Pro Val Phe Arg Asp Asp Val  
405 410 415

Cys Phe Phe Leu Ser Gln Ser Gly Glu Thr Ala Asp Thr Leu Met Gly  
420 425 430

Leu Arg Tyr Cys Lys Glu Arg Gly Ala Leu Thr Val Gly Ile Thr Asn  
435 440 445

Thr Val Gly Ser Ser Ile Ser Arg Glu Thr Asp Cys Gly Val His Ile  
450 455 460

Asn Ala Gly Pro Glu Ile Gly Val Ala Ser Thr Lys Ala Tyr Thr Ser  
 465 470 475 480

Gln Phe Val Ser Leu Val Met Phe Ala Leu Met Met Cys Asp Asp Arg  
 485 490 495

Ile Ser Met Gln Glu Arg Arg Lys Glu Ile Met Leu Gly Leu Lys Arg  
 500 505 510

Leu Pro Asp Leu Ile Lys Glu Val Leu Ser Met Asp Asp Glu Ile Gln  
 515 520 525

Lys Leu Ala Thr Glu Leu Tyr His Gln Lys Ser Val Leu Ile Met Gly  
 530 535 540

Arg Gly Tyr His Tyr Ala Thr Cys Leu Glu Gly Ala Leu Lys Ile Lys  
 545 550 555 560

Glu Ile Thr Tyr Met His Ser Glu Gly Ile Leu Ala Gly Glu Leu Lys  
 565 570 575

His Gly Pro Leu Ala Leu Val Asp Lys Leu Met Pro Val Ile Met Ile  
 580 585 590

Ile Met Arg Asp His Thr Tyr Ala Lys Cys Gln Asn Ala Leu Gln Gln  
 595 600 605

Val Val Ala Arg Gln Gly Arg Pro Val Val Ile Cys Asp Lys Glu Asp  
 610 615 620

Thr Glu Thr Ile Lys Asn Thr Lys Arg Thr Ile Lys Val Pro His Ser  
 625 630 635 640

Val Asp Cys Leu Gln Gly Ile Leu Ser Val Ile Pro Leu Gln Leu Leu  
 645 650 655

Ala Phe His Leu Ala Val Leu Arg Gly Tyr Asp Val Asp Phe Pro Arg  
 660 665 670

Asn Leu Ala Lys Ser Val Thr Val Glu  
 675 680

<210> 3  
 <211> 2049  
 <212> ADN  
 <213> Homo sapiens  
 <220>  
 <221> CDS  
 <222> (1)..(2049)



<223>

<400> 3

atg tgc gga atc ttt gcc tac atg aac tac aga gtc ccc cgg acg agg Met Cys Gly Ile Phe Ala Tyr Met Asn Tyr Arg Val Pro Arg Thr Arg 1 5 10 15	48
aag gag atc ttc gaa acc ctc atc aag ggc ctg cag cgg ctg gag tac Lys Glu Ile Phe Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr 20 25 30	96
aga ggc tac gac tcg gca ggt gtg gcg atc gat ggg aat aat cac gaa Arg Gly Tyr Asp Ser Ala Gly Val Ala Ile Asp Gly Asn Asn His Glu 35 40 45	144
gtc aaa gaa aga cac att cag ctg gtc aag aaa agg ggg aaa gtc aag Val Lys Glu Arg His Ile Gln Leu Val Lys Lys Arg Gly Lys Val Lys 50 55 60	192
gct ctc gat gaa gaa ctt tac aaa caa gac agc atg gac tta aaa gtg Ala Leu Asp Glu Glu Leu Tyr Lys Gln Asp Ser Met Asp Leu Lys Val 65 70 75 80	240
gag ttt gag aca cac ttc ggc att gcc cac acg cgc tgg gcc acc cac Glu Phe Glu Thr His Phe Gly Ile Ala His Thr Arg Trp Ala Thr His 85 90 95	288
ggg gtc ccc agt gct gtc aac agc cac cct cag cgc tca gac aaa ggc Gly Val Pro Ser Ala Val Asn Ser His Pro Gln Arg Ser Asp Lys Gly 100 105 110	336
aac gaa ttt gtt gtc atc cac aat ggg atc atc aca aat tac aaa gat Asn Glu Phe Val Val Ile His Asn Gly Ile Ile Thr Asn Tyr Lys Asp 115 120 125	384
ctg agg aaa ttt ctg gaa agc aaa ggc tac gag ttt gag tca gaa aca Leu Arg Lys Phe Leu Glu Ser Lys Gly Tyr Glu Phe Glu Ser Glu Thr 130 135 140	432
gat aca gag acc atc gcc aag ctg att aaa tat gtg ttc gac aac aga Asp Thr Glu Thr Ile Ala Lys Leu Ile Lys Tyr Val Phe Asp Asn Arg 145 150 155 160	480
gaa act gag gac att acg ttt tca acg ttg gtc gag aga gtc att cag Glu Thr Glu Asp Ile Thr Phe Ser Thr Leu Val Glu Arg Val Ile Gln 165 170 175	528
cag ttg gaa ggt gca ttc gcg ctg gtt ttc aag agt gtc cac tac cca Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val His Tyr Pro 180 185 190	576
gga gaa gcc gtt gcc aca cgg aga ggc agc ccc ctg ctc atc gga gtc Gly Glu Ala Val Ala Thr Arg Arg Gly Ser Pro Leu Leu Ile Gly Val 195 200 205	624
cgg agc aaa tac aag ctc tcc aca gaa cag atc cct atc tta tac agg Arg Ser Lys Tyr Lys Leu Ser Thr Glu Gln Ile Pro Ile Leu Tyr Arg 210 215 220	672
acg tgc act ctg gag aat gtg aag aat atc tgt aag aca cgg atg aag Thr Cys Thr Leu Glu Asn Val Lys Asn Ile Cys Lys Thr Arg Met Lys 225 230 235 240	720
agg ctg gac agc tcc gcc tgc ctg cat gct gtg ggc gac aag gcc gtg Arg Leu Asp Ser Ser Ala Cys Leu His Ala Val Gly Asp Lys Ala Val 245 250 255	768

gaa ttc ttc ttt gct tct gat gca agc gct atc ata gag cac acc aac	816
Glu Phe Phe Phe Ala Ser Asp Ala Ser Ala Ile Ile Glu His Thr Asn	
260 265 270	
cgg gtc atc ttc ctg gag gac gat gac atc gcc gca gtg gct gat ggg	864
Arg Val Ile Phe Leu Glu Asp Asp Ile Ala Ala Val Ala Asp Gly	
275 280 285	
aaa ctc tcc att cac cgg gtc aag cgc tcg gcc agt gat gac cca tct	912
Lys Leu Ser Ile His Arg Val Lys Arg Ser Ala Ser Asp Asp Pro Ser	
290 295 300	
cga gcc atc cag acc ttg cag atg gaa ctg cag caa atc atg aaa ggt	960
Arg Ala Ile Gln Thr Leu Gln Met Glu Leu Gln Gln Ile Met Lys Gly	
305 310 315 320	
aac ttc agt gcg ttt atg cag aag gag atc ttc gaa cag cca gaa tca	1008
Asn Phe Ser Ala Phe Met Gln Lys Glu Ile Phe Glu Gln Pro Glu Ser	
325 330 335	
gtt ttc aat act atg aga ggt cgg gtg aat ttt gaa acc aac aca gtg	1056
Val Phe Asn Thr Met Arg Gly Arg Val Asn Phe Glu Thr Asn Thr Val	
340 345 350	
ctc ctg ggt ggc ttg aag gac cac ttg aag gag att cga cga tgc cga	1104
Leu Leu Gly Gly Leu Lys Asp His Leu Lys Glu Ile Arg Arg Cys Arg	
355 360 365	
cgg ctc atc gtg att ggc tgt gga acc agc tac cac gct gcc gtg gct	1152
Arg Leu Ile Val Ile Gly Cys Gly Thr Ser Tyr His Ala Ala Val Ala	
370 375 380	
acg cgg caa gtt ttg gag gaa ctg act gag ctt cct gtg atg gtt gaa	1200
Thr Arg Gln Val Leu Glu Glu Leu Thr Glu Leu Pro Val Met Val Glu	
385 390 395 400	
ctt gct agt gat ttt ctg gac agg aac aca cct gtg ttc agg gat gac	1248
Leu Ala Ser Asp Phe Leu Asp Arg Asn Thr Pro Val Phe Arg Asp Asp	
405 410 415	
gtt tgc ttt ttc atc agc cag tca ggc gag acc gcg gac acc ctc ctg	1296
Val Cys Phe Phe Ile Ser Gln Ser Gly Glu Thr Ala Asp Thr Leu Leu	
420 425 430	
gcg ctg cgc tac tgt aag gac cgc ggc gct ctc acc gtg ggc gtc acc	1344
Ala Leu Arg Tyr Cys Lys Asp Arg Gly Ala Leu Thr Val Gly Val Thr	
435 440 445	
aac acc gtg ggc agc tcc atc tct cgc gag acc gac tgc ggc gtc cac	1392
Asn Thr Val Gly Ser Ser Ile Ser Arg Glu Thr Asp Cys Gly Val His	
450 455 460	
atc aac gca ggg ccg gag gtc ggc gtg gcc agc acc aag gct tat acc	1440
Ile Asn Ala Gly Pro Glu Val Gly Val Ala Ser Thr Lys Ala Tyr Thr	
465 470 475 480	
agt cag ttc atc tct ctg gtg atg ttt ggt ttg atg atg tct gaa gac	1488
Ser Gln Phe Ile Ser Leu Val Met Phe Gly Leu Met Met Ser Glu Asp	
485 490 495	
cga att tca cta caa aac agg agg caa gag atc atc cgt ggc ttg aga	1536
Arg Ile Ser Leu Gln Asn Arg Arg Gln Glu Ile Ile Arg Gly Leu Arg	
500 505 510	
tct tta cct gag ctg atc aag gaa gtg ctg tct ctg gag gag aag atc	1584

1er dépôt

Ser	Leu	Pro	Glu	Leu	Ile	Lys	Glu	Val	Leu	Ser	Leu	Glu	Glu	Lys	Ile		
515						520						525					
cac	gac	ttg	gcc	ctg	gag	ctc	tac	acg	cag	aga	tcg	ctg	ctg	gtg	atg	1632	
His	Asp	Leu	Ala	Leu	Glu	Leu	Tyr	Thr	Gln	Arg	Ser	Leu	Leu	Val	Met		
530						535						540					
ggg	cgg	ggc	tac	aac	tat	gcc	acc	tgc	ctg	gaa	gga	gcc	ctg	aaa	att	1680	
Gly	Arg	Gly	Tyr	Asn	Tyr	Ala	Thr	Cys	Leu	Glu	Gly	Ala	Leu	Lys	Ile		
545						550						555			560		
aaa	gag	ata	acc	tac	atg	cac	tca	gaa	ggc	atc	ctg	gct	ggg	gag	ctg	1728	
Lys	Glu	Ile	Thr	Tyr	Met	His	Ser	Glu	Gly	Ile	Leu	Ala	Gly	Glu	Leu		
			565						570						575		
aag	cac	ggg	ccc	ctg	gca	ctg	att	gac	aag	cag	atg	ccc	gtc	atc	atg	1776	
Lys	His	Gly	Pro	Leu	Ala	Leu	Ile	Asp	Lys	Gln	Met	Pro	Val	Ile	Met		
			580						585						590		
gtc	att	atg	aag	gat	cct	tgc	ttc	gcc	aaa	tgc	cag	aac	gcc	ctg	cag	1824	
Val	Ile	Met	Lys	Asp	Pro	Cys	Phe	Ala	Lys	Cys	Gln	Asn	Ala	Leu	Gln		
			595						600						605		
caa	gtc	acg	gcc	cgc	cag	ggt	cgc	ccc	att	ata	ctg	tgc	tcc	aag	gac	1872	
Gln	Val	Thr	Ala	Arg	Gln	Gly	Arg	Pro	Ile	Ile	Leu	Cys	Ser	Lys	Asp		
610						615						620					
gat	act	gaa	agt	tcc	aag	ttt	gcg	tat	aag	aca	atc	gag	ctg	ccc	cac	1920	
Asp	Thr	Glu	Ser	Ser	Lys	Phe	Ala	Tyr	Lys	Thr	Ile	Glu	Leu	Pro	His		
625						630						635			640		
act	gtg	gac	tgc	ctc	cag	ggc	atc	ctg	agc	gtg	att	ccg	ctg	cag	ctg	1968	
Thr	Val	Asp	Cys	Leu	Gln	Gly	Ile	Leu	Ser	Val	Ile	Pro	Leu	Gln	Leu		
			645						650						655		
ctg	tcc	ttc	cac	ctg	gct	gtt	ctc	cga	gga	tat	gac	gtt	gac	ttc	ccc	2016	
Leu	Ser	Phe	His	Leu	Ala	Val	Leu	Arg	Gly	Tyr	Asp	Val	Asp	Phe	Pro		
			660						665						670		
aga	aat	ctg	gcc	aag	tct	gta	act	gtg	gaa	tga						2049	
Arg	Asn	Leu	Ala	Lys	Ser	Val	Thr	Val	Glu								
675						680											

<210> 4  
 <211> 682  
 <212> PRT  
 <213> Homo sapiens

<400> 4

Met	Cys	Gly	Ile	Phe	Ala	Tyr	Met	Asn	Tyr	Arg	Val	Pro	Arg	Thr	Arg
1				5					10					15	

Lys	Glu	Ile	Phe	Glu	Thr	Leu	Ile	Lys	Gly	Leu	Gln	Arg	Leu	Glu	Tyr
			20					25					30		

Arg	Gly	Tyr	Asp	Ser	Ala	Gly	Val	Ala	Ile	Asp	Gly	Asn	Asn	His	Glu
		35					40					45			

Val	Lys	Glu	Arg	His	Ile	Gln	Leu	Val	Lys	Lys	Arg	Gly	Lys	Val	Lys
	50					55					60				

1er dépôt

Ala Leu Asp Glu Glu Leu Tyr Lys Gln Asp Ser Met Asp Leu Lys Val  
65 70 75 80

Glu Phe Glu Thr His Phe Gly Ile Ala His Thr Arg Trp Ala Thr His  
85 90 95

Gly Val Pro Ser Ala Val Asn Ser His Pro Gln Arg Ser Asp Lys Gly  
100 105 110

Asn Glu Phe Val Val Ile His Asn Gly Ile Ile Thr Asn Tyr Lys Asp  
115 120 125

Leu Arg Lys Phe Leu Glu Ser Lys Gly Tyr Glu Phe Glu Ser Glu Thr  
130 135 140

Asp Thr Glu Thr Ile Ala Lys Leu Ile Lys Tyr Val Phe Asp Asn Arg  
145 150 155 160

Glu Thr Glu Asp Ile Thr Phe Ser Thr Leu Val Glu Arg Val Ile Gln  
165 170 175

Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val His Tyr Pro  
180 185 190

Gly Glu Ala Val Ala Thr Arg Arg Gly Ser Pro Leu Leu Ile Gly Val  
195 200 205

Arg Ser Lys Tyr Lys Leu Ser Thr Glu Gln Ile Pro Ile Leu Tyr Arg  
210 215 220

Thr Cys Thr Leu Glu Asn Val Lys Asn Ile Cys Lys Thr Arg Met Lys  
225 230 235 240

Arg Leu Asp Ser Ser Ala Cys Leu His Ala Val Gly Asp Lys Ala Val  
245 250 255

Glu Phe Phe Phe Ala Ser Asp Ala Ser Ala Ile Ile Glu His Thr Asn  
260 265 270

Arg Val Ile Phe Leu Glu Asp Asp Asp Ile Ala Ala Val Ala Asp Gly  
275 280 285

Lys Leu Ser Ile His Arg Val Lys Arg Ser Ala Ser Asp Asp Pro Ser  
290 295 300

Arg Ala Ile Gln Thr Leu Gln Met Glu Leu Gln Gln Ile Met Lys Gly  
305 310 315 320

Asn Phe Ser Ala Phe Met Gln Lys Glu Ile Phe Glu Gln Pro Glu Ser

# 1er dépôt

325

330

335

Val Phe Asn Thr Met Arg Gly Arg Val Asn Phe Glu Thr Asn Thr Val  
340 345 350

Leu Leu Gly Gly Leu Lys Asp His Leu Lys Glu Ile Arg Arg Cys Arg  
355 360 365

Arg Leu Ile Val Ile Gly Cys Gly Thr Ser Tyr His Ala Ala Val Ala  
370 375 380

Thr Arg Gln Val Leu Glu Glu Leu Thr Glu Leu Pro Val Met Val Glu  
385 390 395 400

Leu Ala Ser Asp Phe Leu Asp Arg Asn Thr Pro Val Phe Arg Asp Asp  
405 410 415

Val Cys Phe Phe Ile Ser Gln Ser Gly Glu Thr Ala Asp Thr Leu Leu  
420 425 430

Ala Leu Arg Tyr Cys Lys Asp Arg Gly Ala Leu Thr Val Gly Val Thr  
435 440 445

Asn Thr Val Gly Ser Ser Ile Ser Arg Glu Thr Asp Cys Gly Val His  
450 455 460

Ile Asn Ala Gly Pro Glu Val Gly Val Ala Ser Thr Lys Ala Tyr Thr  
465 470 475 480

Ser Gln Phe Ile Ser Leu Val Met Phe Gly Leu Met Met Ser Glu Asp  
485 490 495

Arg Ile Ser Leu Gln Asn Arg Arg Gln Glu Ile Ile Arg Gly Leu Arg  
500 505 510

Ser Leu Pro Glu Leu Ile Lys Glu Val Leu Ser Leu Glu Glu Lys Ile  
515 520 525

His Asp Leu Ala Leu Glu Leu Tyr Thr Gln Arg Ser Leu Leu Val Met  
530 535 540

Gly Arg Gly Tyr Asn Tyr Ala Thr Cys Leu Glu Gly Ala Leu Lys Ile  
545 550 555 560

Lys Glu Ile Thr Tyr Met His Ser Glu Gly Ile Leu Ala Gly Glu Leu  
565 570 575

Lys His Gly Pro Leu Ala Leu Ile Asp Lys Gln Met Pro Val Ile Met  
580 585 590

Val Ile Met Lys Asp Pro Cys Phe Ala Lys Cys Gln Asn Ala Leu Gln  
595 600 605

Gln Val Thr Ala Arg Gln Gly Arg Pro Ile Ile Leu Cys Ser Lys Asp  
610 615 620

Asp Thr Glu Ser Ser Lys Phe Ala Tyr Lys Thr Ile Glu Leu Pro His  
625 630 635 640

Thr Val Asp Cys Leu Gln Gly Ile Leu Ser Val Ile Pro Leu Gln Leu  
645 650 655

Leu Ser Phe His Leu Ala Val Leu Arg Gly Tyr Asp Val Asp Phe Pro  
660 665 670

Arg Asn Leu Ala Lys Ser Val Thr Val Glu  
675 680

<210> 5  
<211> 2100  
<212> ADN  
<213> Homo sapiens

<220>  
<221> CDS  
<222> (1)..(2100)  
<223>

<220>  
<221> misc feature  
<222> (170)..(170)  
<223> t ou c

<400> 5  
atg tgt ggt ata ttt gct tac tta aac tac cat gtt cct cga acg aga 48  
Met Cys Gly Ile Phe Ala Tyr Leu Asn Tyr His Val Pro Arg Thr Arg  
1 5 10 15

cga gaa atc ctg gag acc cta atc aaa ggc ctt cag aga ctg gag tac 96  
Arg Glu Ile Leu Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr  
20 25 30

aga gga tat gat tct gct ggt gtg gga ttt gat gga ggc aat gat aaa 144  
Arg Gly Tyr Asp Ser Ala Gly Val Gly Phe Asp Gly Gly Asn Asp Lys  
35 40 45

gat tgg gaa gcc aat gcc tgc aaa anc cag ctt att aag aag aaa gga 192  
Asp Trp Glu Ala Asn Ala Cys Lys Xaa Gln Leu Ile Lys Lys Lys Gly  
50 55 60

aaa gtt aag gca ctg gat gaa gaa gtt cac aag caa caa gat atg gat 240  
Lys Val Lys Ala Leu Asp Glu Glu Val His Lys Gln Gln Asp Met Asp  
65 70 75 80

ttg gat ata gaa ttt gat gta cac ctt gga ata gct cat acc cgt tgg 288  
Leu Asp Ile Glu Phe Asp Val His Leu Gly Ile Ala His Thr Arg Trp  
85 90 95

gca aca cat gga gaa ccc agt cct gtc aat agc cac ccc cag cgc tct Ala Thr His Gly Glu Pro Ser Pro Val Asn Ser His Pro Gln Arg Ser 100 105 110	336
gat aaa aat aat gaa ttt atc gtt att cac aat gga atc atc acc aac Asp Lys Asn Asn Glu Phe Ile Val Ile His Asn Gly Ile Ile Thr Asn 115 120 125	384
tac aaa gac ttg aaa aag ttt ttg gaa agc aaa ggc tat gac ttc gaa Tyr Lys Asp Leu Lys Lys Phe Leu Glu Ser Lys Gly Tyr Asp Phe Glu 130 135 140	432
tct gaa aca gac aca gag aca att gcc aag ctc gtt aag tat atg tat Ser Glu Thr Asp Thr Glu Thr Ile Ala Lys Leu Val Lys Tyr Met Tyr 145 150 155 160	480
gac aat cgg gaa agt caa gat acc agc ttt act acc ttg gtg gag aga Asp Asn Arg Glu Ser Gln Asp Thr Ser Phe Thr Thr Leu Val Glu Arg 165 170 175	528
gtt atc caa caa ttg gaa ggt gct ttt gca ctt gtg ttt aaa agt gtt Val Ile Gln Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val 180 185 190	576
cat ttt ccc ggg caa gca gtt ggc aca agg cga ggt agc cct ctg ttg His Phe Pro Gly Gln Ala Val Gly Thr Arg Arg Gly Ser Pro Leu Leu 195 200 205	624
att ggt gta cgg agt gaa cat aaa ctt tct act gat cac att cct ata Ile Gly Val Arg Ser Glu His Lys Leu Ser Thr Asp His Ile Pro Ile 210 215 220	672
ctc tac aga aca gct agg act cag att gga tca aaa ttc aca cgg tgg Leu Tyr Arg Thr Ala Arg Thr Gln Ile Gly Ser Lys Phe Thr Arg Trp 225 230 235 240	720
gga tca cag gga gaa aga ggc aaa gac aag aaa gga agc tgc aat ctc Gly Ser Gln Gly Glu Arg Gly Lys Asp Lys Lys Gly Ser Cys Asn Leu 245 250 255	768
tct cgt gtg gac agc aca acc tgc ctt ttc ccg gtg gaa gaa aaa gca Ser Arg Val Asp Ser Thr Thr Cys Leu Phe Pro Val Glu Glu Lys Ala 260 265 270	816
gtg gag tat tac ttt gct tct gat gca agt gct gtc ata gaa cac acc Val Glu Tyr Tyr Phe Ala Ser Asp Ala Ser Ala Val Ile Glu His Thr 275 280 285	864
aat cgc gtc atc ttt ctg gaa gat gat gat gtt gca gca gta gtg gat Asn Arg Val Ile Phe Leu Glu Asp Asp Asp Val Ala Ala Val Val Asp 290 295 300	912
gga cgt ctt tct atc cat cga att aaa cga act gca gga gat cac ccc Gly Arg Leu Ser Ile His Arg Ile Lys Arg Thr Ala Gly Asp His Pro 305 310 315 320	960
gga cga gct gtg caa aca ctc cag atg gaa ctc cag cag atc atg aag Gly Arg Ala Val Gln Thr Leu Gln Met Glu Leu Gln Gln Ile Met Lys 325 330 335	1008
ggc aac ttc agt tca ttt atg cag aag gaa ata ttt gag cag cca gag Gly Asn Phe Ser Ser Phe Met Gln Lys Glu Ile Phe Glu Gln Pro Glu 340 345 350	1056
tct gtc gtg aac aca atg aga gga aga gtc aac ttt gat gac tat act Ser Val Val Asn Thr Met Arg Gly Arg Val Asn Phe Asp Asp Tyr Thr 350	1104

355	360	365	
gtg aat ttg ggt ggt ttg aag gat cac ata aag gag atc cag aga tgc Val Asn Leu Gly Gly Leu Lys Asp His Ile Lys Glu Ile Gln Arg Cys 370 375 380			1152
cgg cgt ttg att ctt att gct tgt gga aca agt tac cat gct ggt gta Arg Arg Leu Ile Leu Ile Ala Cys Gly Thr Ser Tyr His Ala Gly Val 385 390 395 400			1200
gca aca cgt caa gtt ctt gag gag ctg act gag ttg cct gtg atg gtg Ala Thr Arg Gln Val Leu Glu Glu Leu Thr Glu Leu Pro Val Met Val 405 410 415			1248
gaa cta gca agt gac ttc ctg gac aga aac aca cca gtc ttt cga gat Glu Leu Ala Ser Asp Phe Leu Asp Arg Asn Thr Pro Val Phe Arg Asp 420 425 430			1296
gat gtt tgc ttt ttc ctt agt caa tca ggt gag aca gca gat act ttg Asp Val Cys Phe Phe Leu Ser Gln Ser Gly Glu Thr Ala Asp Thr Leu 435 440 445			1344
atg ggt ctt cgt tac tgt aag gag aga gga gct tta act gtg ggg atc Met Gly Leu Arg Tyr Cys Lys Glu Arg Gly Ala Leu Thr Val Gly Ile 450 455 460			1392
aca aac aca gtt ggc agt tcc ata tca cgg gag aca gat tgt gga gtt Thr Asn Thr Val Gly Ser Ser Ile Ser Arg Glu Thr Asp Cys Gly Val 465 470 475 480			1440
cat att aat gct ggt cct gag att ggt gtg gcc agt aca aag gct tat His Ile Asn Ala Gly Pro Glu Ile Gly Val Ala Ser Thr Lys Ala Tyr 485 490 495			1488
acc agc cag ttt gta tcc ctt gtg atg ttt gcc ctt atg atg tgt gat Thr Ser Gln Phe Val Ser Leu Val Met Phe Ala Leu Met Met Cys Asp 500 505 510			1536
gat cgg atc tcc atg caa gaa aga cgc aaa gag atc atg ctt gga ttg Asp Arg Ile Ser Met Gln Glu Arg Arg Lys Glu Ile Met Leu Gly Leu 515 520 525			1584
aaa cgg ctg cct gat ttg att aag gaa gta ctg agc atg gat gac gaa Lys Arg Leu Pro Asp Leu Ile Lys Glu Val Leu Ser Met Asp Asp Glu 530 535 540			1632
att cag aaa cta gca aca gaa ctt tat cat cag aag tca gtt ctg ata Ile Gln Lys Leu Ala Thr Glu Leu Tyr His Gln Lys Ser Val Leu Ile 545 550 555 560			1680
atg gga cga ggc tat cat tat gct act tgt ctt gaa ggg gca ctg aaa Met Gly Arg Gly Tyr His Tyr Ala Thr Cys Leu Glu Gly Ala Leu Lys 565 570 575			1728
atc aaa gaa att act tat atg cac tct gaa ggc atc ctt gct ggt gaa Ile Lys Glu Ile Thr Tyr Met His Ser Glu Gly Ile Leu Ala Gly Glu 580 585 590			1776
ttg aaa cat ggc cct ctg gct ttg gtg gat aaa ttg atg cct gtg atc Leu Lys His Gly Pro Leu Ala Leu Val Asp Lys Leu Met Pro Val Ile 595 600 605			1824
atg atc atc atg aga gat cac act tat gcc aag tgt cag aat gct ctt Met Ile Ile Met Arg Asp His Thr Tyr Ala Lys Cys Gln Asn Ala Leu 610 615 620			1872



# 1er dépôt

cag caa gtg gtt gct cgg cag ggg cgg cct gtg gta att tgt gat aag 1920  
Gln Gln Val Val Ala Arg Gln Gly Arg Pro Val Val Ile Cys Asp Lys  
625 630 635 640

gag gat act gag acc att aag aac aca aaa aga acg atc aag gtg ccc 1968  
Glu Asp Thr Glu Thr Ile Lys Asn Thr Lys Arg Thr Ile Lys Val Pro  
645 650 655

cac tca gtg gac tgc ttg cag ggc att ctc agc gtg atc cct tta cag 2016  
His Ser Val Asp Cys Leu Gln Gly Ile Leu Ser Val Ile Pro Leu Gln  
660 665 670

ttg ctg gct ttc cac ctt gct gtg ctg aga ggc tat gat gtt gat ttc 2064  
Leu Leu Ala Phe His Leu Ala Val Leu Arg Gly Tyr Asp Val Asp Phe  
675 680 685

cca cgg aat ctt gcc aaa tct gtg act gta gag tga 2100  
Pro Arg Asn Leu Ala Lys Ser Val Thr Val Glu  
690 695

<210> 6  
<211> 699  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<222> (57)..(57)  
<223> Le 'Xaa' en position 57 représente Thr ou Ile.

<400> 6

Met Cys Gly Ile Phe Ala Tyr Leu Asn Tyr His Val Pro Arg Thr Arg  
1 5 10 15

Arg Glu Ile Leu Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr  
20 25 30

Arg Gly Tyr Asp Ser Ala Gly Val Gly Phe Asp Gly Gly Asn Asp Lys  
35 40 45

Asp Trp Glu Ala Asn Ala Cys Lys Xaa Gln Leu Ile Lys Lys Lys Gly  
50 55 60

Lys Val Lys Ala Leu Asp Glu Glu Val His Lys Gln Gln Asp Met Asp  
65 70 75 80

Leu Asp Ile Glu Phe Asp Val His Leu Gly Ile Ala His Thr Arg Trp  
85 90 95

Ala Thr His Gly Glu Pro Ser Pro Val Asn Ser His Pro Gln Arg Ser  
100 105 110

Asp Lys Asn Asn Glu Phe Ile Val Ile His Asn Gly Ile Ile Thr Asn  
115 120 125

Tyr Lys Asp Leu Lys Lys Phe Leu Glu Ser Lys Gly Tyr Asp Phe Glu

# 1er dépôt

130

135

140

Ser Glu Thr Asp Thr Glu Thr Ile Ala Lys Leu Val Lys Tyr Met Tyr  
145 150 155 160

Asp Asn Arg Glu Ser Gln Asp Thr Ser Phe Thr Thr Leu Val Glu Arg  
165 170 175

Val Ile Gln Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val  
180 185 190

His Phe Pro Gly Gln Ala Val Gly Thr Arg Arg Gly Ser Pro Leu Leu  
195 200 205

Ile Gly Val Arg Ser Glu His Lys Leu Ser Thr Asp His Ile Pro Ile  
210 215 220

Leu Tyr Arg Thr Ala Arg Thr Gln Ile Gly Ser Lys Phe Thr Arg Trp  
225 230 235 240

Gly Ser Gln Gly Glu Arg Gly Lys Asp Lys Lys Gly Ser Cys Asn Leu  
245 250 255

Ser Arg Val Asp Ser Thr Thr Cys Leu Phe Pro Val Glu Glu Lys Ala  
260 265 270

Val Glu Tyr Tyr Phe Ala Ser Asp Ala Ser Ala Val Ile Glu His Thr  
275 280 285

Asn Arg Val Ile Phe Leu Glu Asp Asp Asp Val Ala Ala Val Val Asp  
290 295 300

Gly Arg Leu Ser Ile His Arg Ile Lys Arg Thr Ala Gly Asp His Pro  
305 310 315 320

Gly Arg Ala Val Gln Thr Leu Gln Met Glu Leu Gln Gln Ile Met Lys  
325 330 335

Gly Asn Phe Ser Ser Phe Met Gln Lys Glu Ile Phe Glu Gln Pro Glu  
340 345 350

Ser Val Val Asn Thr Met Arg Gly Arg Val Asn Phe Asp Asp Tyr Thr  
355 360 365

Val Asn Leu Gly Gly Leu Lys Asp His Ile Lys Glu Ile Gln Arg Cys  
370 375 380

Arg Arg Leu Ile Leu Ile Ala Cys Gly Thr Ser Tyr His Ala Gly Val  
385 390 395 400

Ala Thr Arg Gln Val Leu Glu Glu Leu Thr Glu Leu Pro Val Met Val  
405 410 415

Glu Leu Ala Ser Asp Phe Leu Asp Arg Asn Thr Pro Val Phe Arg Asp  
420 425 430

Asp Val Cys Phe Phe Leu Ser Gln Ser Gly Glu Thr Ala Asp Thr Leu  
435 440 445

Met Gly Leu Arg Tyr Cys Lys Glu Arg Gly Ala Leu Thr Val Gly Ile  
450 455 460

Thr Asn Thr Val Gly Ser Ser Ile Ser Arg Glu Thr Asp Cys Gly Val  
465 470 475 480

His Ile Asn Ala Gly Pro Glu Ile Gly Val Ala Ser Thr Lys Ala Tyr  
485 490 495

Thr Ser Gln Phe Val Ser Leu Val Met Phe Ala Leu Met Met Cys Asp  
500 505 510

Asp Arg Ile Ser Met Gln Glu Arg Arg Lys Glu Ile Met Leu Gly Leu  
515 520 525

Lys Arg Leu Pro Asp Leu Ile Lys Glu Val Leu Ser Met Asp Asp Glu  
530 535 540

Ile Gln Lys Leu Ala Thr Glu Leu Tyr His Gln Lys Ser Val Leu Ile  
545 550 555 560

Met Gly Arg Gly Tyr His Tyr Ala Thr Cys Leu Glu Gly Ala Leu Lys  
565 570 575

Ile Lys Glu Ile Thr Tyr Met His Ser Glu Gly Ile Leu Ala Gly Glu  
580 585 590

Leu Lys His Gly Pro Leu Ala Leu Val Asp Lys Leu Met Pro Val Ile  
595 600 605

Met Ile Ile Met Arg Asp His Thr Tyr Ala Lys Cys Gln Asn Ala Leu  
610 615 620

Gln Gln Val Val Ala Arg Gln Gly Arg Pro Val Val Ile Cys Asp Lys  
625 630 635 640

Glu Asp Thr Glu Thr Ile Lys Asn Thr Lys Arg Thr Ile Lys Val Pro  
645 650 655

His Ser Val Asp Cys Leu Gln Gly Ile Leu Ser Val Ile Pro Leu Gln

1er dépôt

660

665

670

Leu Leu Ala Phe His Leu Ala Val Leu Arg Gly Tyr Asp Val Asp Phe  
675 680 685

Pro Arg Asn Leu Ala Lys Ser Val Thr Val Glu  
690 695

<210> 7  
<211> 2064  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> GFAT1 modifiée par une étiquette de purification interne

<220>  
<221> CDS  
<222> (1)..(2064)  
<223>

<220>  
<221> misc\_feature  
<222> (170)..(170)  
<223> t ou c

<400> 7  
atg tgt ggt ata ttt gct tac tta aac tac cat gtt cct cga acg aga 48  
Met Cys Gly Ile Phe Ala Tyr Leu Asn Tyr His Val Pro Arg Thr Arg  
1 5 10 15  
cga gaa atc ctg gag acc cta atc aaa ggc ctt cag aga ctg gag tac 96  
Arg Glu Ile Leu Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr  
20 25 30  
aga gga tat gat tct gct ggt gtg gga ttt gat gga ggc aat gat aaa 144  
Arg Gly Tyr Asp Ser Ala Gly Val Gly Phe Asp Gly Gly Asn Asp Lys  
35 40 45  
gat tgg gaa gcc aat gcc tgc aaa anc cag ctt att aag aag aaa gga 192  
Asp Trp Glu Ala Asn Ala Cys Lys Xaa Gln Leu Ile Lys Lys Lys Gly  
50 55 60  
aaa gtt aag gca ctg gat gaa gaa gtt cac aag caa caa gat atg gat 240  
Lys Val Lys Ala Leu Asp Glu Glu Val His Lys Gln Gln Asp Met Asp  
65 70 75 80  
ttg gat ata gaa ttt gat gta cac ctt gga ata gct cat acc cgt tgg 288  
Leu Asp Ile Glu Phe Asp Val His Leu Gly Ile Ala His Thr Arg Trp  
85 90 95  
gca aca cat gga gaa ccc agt cct gtc aat agc cac ccc cag cgc tct 336  
Ala Thr His Gly Glu Pro Ser Pro Val Asn Ser His Pro Gln Arg Ser  
100 105 110  
gat aaa aat aat gaa ttt atc gtt att cac aat gga atc atc acc aac 384  
Asp Lys Asn Asn Glu Phe Ile Val Ile His Asn Gly Ile Ile Thr Asn  
115 120 125  
tac aaa gac ttg aaa aag ttt ttg gaa agc aaa ggc tat gac ttc gaa 432  
Tyr Lys Asp Leu Lys Lys Phe Leu Glu Ser Lys Gly Tyr Asp Phe Glu  
130 135 140

1er dépôt

tct gaa aca gac aca gag aca att gcc aag ctc gtt aag tat atg tat Ser Glu Thr Asp Thr Glu Thr Ile Ala Lys Leu Val Lys Tyr Met Tyr 145 150 155 160	480
gac aat cgg gaa agt caa gat acc agc ttt act acc ttg gtg gag aga Asp Asn Arg Glu Ser Gln Asp Thr Ser Phe Thr Thr Leu Val Glu Arg 165 170 175	528
gtt atc caa caa ttg gaa ggt gct ttt gca ctt gtg ttt aaa agt gtt Val Ile Gln Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val 180 185 190	576
cat ttt ccc ggg caa gca gtt ggc aca agg cga ggt agc cct ctg ttg His Phe Pro Gly Gln Ala Val Gly Thr Arg Arg Gly Ser Pro Leu Leu 195 200 205	624
att ggt gta cgg agt gaa cat aaa ctt tct act gat cac att cct ata Ile Gly Val Arg Ser Glu His Lys Leu Ser Thr Asp His Ile Pro Ile 210 215 220	672
ctc tac aga aca ggc aaa gac aag aaa gga agc tgc aat ctc tct cgt Leu Tyr Arg Thr Gly Lys Asp Lys Lys Gly Ser Cys Asn Leu Ser Arg 225 230 235 240	720
gtg gac agc aca acc tgc ctt ttc ccg gtg gaa gaa aaa gca gtg gag Val Asp Ser Thr Thr Cys Leu Phe Pro Val Glu Glu Lys Ala Val Glu 245 250 255	768
tat tac ttt gct tct gat gca agt gct gtc ata gaa cac acc aat cgc Tyr Tyr Phe Ala Ser Asp Ala Ser Ala Val Ile Glu His Thr Asn Arg 260 265 270	816
gtc atc ttt ctg gaa gat gat gat gtt gca gca gta gtg gat gga cgt Val Ile Phe Leu Glu Asp Asp Asp Val Ala Ala Val Val Asp Gly Arg 275 280 285	864
ctt tct atc cat cga att aaa cga act gca gga cat cac cat cac cat Leu Ser Ile His Arg Ile Lys Arg Thr Ala Gly His His His His His 290 295 300	912
cac gat cac ccc gga cga gct gtg caa aca ctc cag atg gaa ctc cag His Asp His Pro Gly Arg Ala Val Gln Thr Leu Gln Met Glu Leu Gln 305 310 315 320	960
cag atc atg aag ggc aac ttc agt tca ttt atg cag aag gaa ata ttt Gln Ile Met Lys Gly Asn Phe Ser Ser Phe Met Gln Lys Glu Ile Phe 325 330 335	1008
gag cag cca gag tct gtc gtg aac aca atg aga gga aga gtc aac ttt Glu Gln Pro Glu Ser Val Val Asn Thr Met Arg Gly Arg Val Asn Phe 340 345 350	1056
gat gac tat act gtg aat ttg ggt ggt ttg aag gat cac ata aag gag Asp Asp Tyr Thr Val Asn Leu Gly Gly Leu Lys Asp His Ile Lys Glu 355 360 365	1104
atc cag aga tgc cgg cgt ttg att ctt att gct tgt gga aca agt tac Ile Gln Arg Cys Arg Arg Leu Ile Leu Ile Ala Cys Gly Thr Ser Tyr 370 375 380	1152
cat gct ggt gta gca aca cgt caa gtt ctt gag gag ctg act gag ttg His Ala Gly Val Ala Thr Arg Gln Val Leu Glu Glu Leu Thr Glu Leu 385 390 395 400	1200
cct gtg atg gtg gaa cta gca agt gac ttc ctg gac aga aac aca cca	1248

Pro	Val	Met	Val	Glu	Leu	Ala	Ser	Asp	Phe	Leu	Asp	Arg	Asn	Thr	Pro	
				405					410					415		
gtc	ttt	cga	gat	gat	gtt	tgc	ttt	ttc	ctt	agt	caa	tca	ggt	gag	aca	1296
Val	Phe	Arg	Asp	Asp	Val	Cys	Phe	Phe	Leu	Ser	Gln	Ser	Gly	Glu	Thr	
			420					425					430			
gca	gat	act	ttg	atg	ggt	ctt	cgt	tac	tgt	aag	gag	aga	gga	gct	tta	1344
Ala	Asp	Thr	Leu	Met	Gly	Leu	Arg	Tyr	Cys	Lys	Glu	Arg	Gly	Ala	Leu	
			435				440					445				
act	gtg	ggg	atc	aca	aac	aca	gtt	ggc	agt	tcc	ata	tca	cgg	gag	aca	1392
Thr	Val	Gly	Ile	Thr	Asn	Thr	Val	Gly	Ser	Ser	Ile	Ser	Arg	Glu	Thr	
	450					455					460					
gat	tgt	gga	gtt	cat	att	aat	gct	ggt	cct	gag	att	ggt	gtg	gcc	agt	1440
Asp	Cys	Gly	Val	His	Ile	Asn	Ala	Gly	Pro	Glu	Ile	Gly	Val	Ala	Ser	
465					470				475						480	
aca	aag	gct	tat	acc	agc	cag	ttt	gta	tcc	ctt	gtg	atg	ttt	gcc	ctt	1488
Thr	Lys	Ala	Tyr	Thr	Ser	Gln	Phe	Val	Ser	Leu	Val	Met	Phe	Ala	Leu	
				485					490					495		
atg	atg	tgt	gat	gat	cgg	atc	tcc	atg	caa	gaa	aga	cgc	aaa	gag	atc	1536
Met	Met	Cys	Asp	Asp	Arg	Ile	Ser	Met	Gln	Glu	Arg	Arg	Lys	Glu	Ile	
			500					505					510			
atg	ctt	gga	ttg	aaa	cgg	ctg	cct	gat	ttg	att	aag	gaa	gta	ctg	agc	1584
Met	Leu	Gly	Leu	Lys	Arg	Leu	Pro	Asp	Leu	Ile	Lys	Glu	Val	Leu	Ser	
		515					520					525				
atg	gat	gac	gaa	att	cag	aaa	cta	gca	aca	gaa	ctt	tat	cat	cag	aag	1632
Met	Asp	Asp	Glu	Ile	Gln	Lys	Leu	Ala	Thr	Glu	Leu	Tyr	His	Gln	Lys	
	530					535					540					
tca	gtt	ctg	ata	atg	gga	cga	ggc	tat	cat	tat	gct	act	tgt	ctt	gaa	1680
Ser	Val	Leu	Ile	Met	Gly	Arg	Gly	Tyr	His	Tyr	Ala	Thr	Cys	Leu	Glu	
545					550				555						560	
ggg	gca	ctg	aaa	atc	aaa	gaa	att	act	tat	atg	cac	tct	gaa	ggc	atc	1728
Gly	Ala	Leu	Lys	Ile	Lys	Glu	Ile	Thr	Tyr	Met	His	Ser	Glu	Gly	Ile	
			565						570					575		
ctt	gct	ggt	gaa	ttg	aaa	cat	ggc	cct	ctg	gct	ttg	gtg	gat	aaa	ttg	1776
Leu	Ala	Gly	Glu	Leu	Lys	His	Gly	Pro	Leu	Ala	Leu	Val	Asp	Lys	Leu	
			580					585					590			
atg	cct	gtg	atc	atg	atc	atc	atg	aga	gat	cac	act	tat	gcc	aag	tgt	1824
Met	Pro	Val	Ile	Met	Ile	Ile	Met	Arg	Asp	His	Thr	Tyr	Ala	Lys	Cys	
		595					600					605				
cag	aat	gct	ctt	cag	caa	gtg	gtt	gct	cgg	cag	ggg	cgg	cct	gtg	gta	1872
Gln	Asn	Ala	Leu	Gln	Gln	Val	Val	Ala	Arg	Gln	Gly	Arg	Pro	Val	Val	
			610			615					620					
att	tgt	gat	aag	gag	gat	act	gag	acc	att	aag	aac	aca	aaa	aga	acg	1920
Ile	Cys	Asp	Lys	Glu	Asp	Thr	Glu	Thr	Ile	Lys	Asn	Thr	Lys	Arg	Thr	
625					630					635					640	
atc	aag	gtg	ccc	cac	tca	gtg	gac	tgc	ttg	cag	ggc	att	ctc	agc	gtg	1968
Ile	Lys	Val	Pro	His	Ser	Val	Asp	Cys	Leu	Gln	Gly	Ile	Leu	Ser	Val	
				645					650					655		
atc	cct	tta	cag	ttg	ctg	gct	ttc	cac	ctt	gct	gtg	ctg	aga	ggc	tat	2016
Ile	Pro	Leu	Gln	Leu	Leu	Ala	Phe	His	Leu	Ala	Val	Leu	Arg	Gly	Tyr	
			660					665							670	

gat gtt gat ttc cca cgg aat ctt gcc aaa tct gtg act gta gag tga 2064  
 . Asp Val Asp Phe Pro Arg Asn Leu Ala Lys Ser Val Thr Val Glu  
 675 680 685

<210> 8  
 <211> 687  
 <212> PRT  
 <213> Séquence artificielle

<220>  
 <221> misc\_feature  
 <222> (57)..(57)  
 <223> Le 'Xaa' en position 57 représente Thr ou Ile.  
 <220>  
 <223> GFAT1 modifiée par une étiquette de purification interne  
 <400> 8

Met Cys Gly Ile Phe Ala Tyr Leu Asn Tyr His Val Pro Arg Thr Arg  
 1 5 10 15

Arg Glu Ile Leu Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr  
 20 25 30

Arg Gly Tyr Asp Ser Ala Gly Val Gly Phe Asp Gly Gly Asn Asp Lys  
 35 40 45

Asp Trp Glu Ala Asn Ala Cys Lys Xaa Gln Leu Ile Lys Lys Lys Gly  
 50 55 60

Lys Val Lys Ala Leu Asp Glu Glu Val His Lys Gln Gln Asp Met Asp  
 65 70 75 80

Leu Asp Ile Glu Phe Asp Val His Leu Gly Ile Ala His Thr Arg Trp  
 85 90 95

Ala Thr His Gly Glu Pro Ser Pro Val Asn Ser His Pro Gln Arg Ser  
 100 105 110

Asp Lys Asn Asn Glu Phe Ile Val Ile His Asn Gly Ile Ile Thr Asn  
 115 120 125

Tyr Lys Asp Leu Lys Lys Phe Leu Glu Ser Lys Gly Tyr Asp Phe Glu  
 130 135 140

Ser Glu Thr Asp Thr Glu Thr Ile Ala Lys Leu Val Lys Tyr Met Tyr  
 145 150 155 160

Asp Asn Arg Glu Ser Gln Asp Thr Ser Phe Thr Thr Leu Val Glu Arg  
 165 170 175

Val Ile Gln Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val

180

185

190

His Phe Pro Gly Gln Ala Val Gly Thr Arg Arg Gly Ser Pro Leu Leu  
195 200 205

Ile Gly Val Arg Ser Glu His Lys Leu Ser Thr Asp His Ile Pro Ile  
210 215 220

Leu Tyr Arg Thr Gly Lys Asp Lys Lys Gly Ser Cys Asn Leu Ser Arg  
225 230 235 240

Val Asp Ser Thr Thr Cys Leu Phe Pro Val Glu Glu Lys Ala Val Glu  
245 250 255

Tyr Tyr Phe Ala Ser Asp Ala Ser Ala Val Ile Glu His Thr Asn Arg  
260 265 270

Val Ile Phe Leu Glu Asp Asp Asp Val Ala Ala Val Val Asp Gly Arg  
275 280 285

Leu Ser Ile His Arg Ile Lys Arg Thr Ala Gly His His His His His  
290 295 300

His Asp His Pro Gly Arg Ala Val Gln Thr Leu Gln Met Glu Leu Gln  
305 310 315 320

Gln Ile Met Lys Gly Asn Phe Ser Ser Phe Met Gln Lys Glu Ile Phe  
325 330 335

Glu Gln Pro Glu Ser Val Val Asn Thr Met Arg Gly Arg Val Asn Phe  
340 345 350

Asp Asp Tyr Thr Val Asn Leu Gly Gly Leu Lys Asp His Ile Lys Glu  
355 360 365

Ile Gln Arg Cys Arg Arg Leu Ile Leu Ile Ala Cys Gly Thr Ser Tyr  
370 375 380

His Ala Gly Val Ala Thr Arg Gln Val Leu Glu Glu Leu Thr Glu Leu  
385 390 395 400

Pro Val Met Val Glu Leu Ala Ser Asp Phe Leu Asp Arg Asn Thr Pro  
405 410 415

Val Phe Arg Asp Asp Val Cys Phe Phe Leu Ser Gln Ser Gly Glu Thr  
420 425 430

Ala Asp Thr Leu Met Gly Leu Arg Tyr Cys Lys Glu Arg Gly Ala Leu  
435 440 445



Thr Val Gly Ile Thr Asn Thr Val Gly Ser Ser Ile Ser Arg Glu Thr  
 450 455 460

Asp Cys Gly Val His Ile Asn Ala Gly Pro Glu Ile Gly Val Ala Ser  
 465 470 475 480

Thr Lys Ala Tyr Thr Ser Gln Phe Val Ser Leu Val Met Phe Ala Leu  
 485 490 495

Met Met Cys Asp Asp Arg Ile Ser Met Gln Glu Arg Arg Lys Glu Ile  
 500 505 510

Met Leu Gly Leu Lys Arg Leu Pro Asp Leu Ile Lys Glu Val Leu Ser  
 515 520 525

Met Asp Asp Glu Ile Gln Lys Leu Ala Thr Glu Leu Tyr His Gln Lys  
 530 535 540

Ser Val Leu Ile Met Gly Arg Gly Tyr His Tyr Ala Thr Cys Leu Glu  
 545 550 555 560

Gly Ala Leu Lys Ile Lys Glu Ile Thr Tyr Met His Ser Glu Gly Ile  
 565 570 575

Leu Ala Gly Glu Leu Lys His Gly Pro Leu Ala Leu Val Asp Lys Leu  
 580 585 590

Met Pro Val Ile Met Ile Ile Met Arg Asp His Thr Tyr Ala Lys Cys  
 595 600 605

Gln Asn Ala Leu Gln Gln Val Val Ala Arg Gln Gly Arg Pro Val Val  
 610 615 620

Ile Cys Asp Lys Glu Asp Thr Glu Thr Ile Lys Asn Thr Lys Arg Thr  
 625 630 635 640

Ile Lys Val Pro His Ser Val Asp Cys Leu Gln Gly Ile Leu Ser Val  
 645 650 655

Ile Pro Leu Gln Leu Leu Ala Phe His Leu Ala Val Leu Arg Gly Tyr  
 660 665 670

Asp Val Asp Phe Pro Arg Asn Leu Ala Lys Ser Val Thr Val Glu  
 675 680 685

<210> 9  
 <211> 2067  
 <212> ADN  
 <213> Séquence artificielle

```

<220>
<223> GFAT2 modifiée par une étiquette de purification interne

<220>
<221> CDS
<222> (1)..(2067)
<223>

<400> 9
atg tgc gga atc ttt gcc tac atg aac tac aga gtc ccc cgg acg agg      48
Met Cys Gly Ile Phe Ala Tyr Met Asn Tyr Arg Val Pro Arg Thr Arg
1          5          10          15

aag gag atc ttc gaa acc ctc atc aag ggc ctg cag cgg ctg gag tac      96
Lys Glu Ile Phe Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr
          20          25          30

aga ggc tac gac tcg gca ggt gtg gcg atc gat ggg aat aat cac gaa      144
Arg Gly Tyr Asp Ser Ala Gly Val Ala Ile Asp Gly Asn Asn His Glu
          35          40          45

gtc aaa gaa aga cac att cag ctg gtc aag aaa agg ggg aaa gtc aag      192
Val Lys Glu Arg His Ile Gln Leu Val Lys Lys Arg Gly Lys Val Lys
          50          55          60

gct ctc gat gaa gaa ctt tac aaa caa gac agc atg gac tta aaa gtg      240
Ala Leu Asp Glu Glu Leu Tyr Lys Gln Asp Ser Met Asp Leu Lys Val
65          70          75          80

gag ttt gag aca cac ttc ggc att gcc cac acg cgc tgg gcc acc cac      288
Glu Phe Glu Thr His Phe Gly Ile Ala His Thr Arg Trp Ala Thr His
          85          90          95

ggg gtc ccc agt gct gtc aac agc cac cct cag cgc tca gac aaa ggc      336
Gly Val Pro Ser Ala Val Asn Ser His Pro Gln Arg Ser Asp Lys Gly
          100          105          110

aac gaa ttt gtt gtc atc cac aat ggg atc atc aca aat tac aaa gat      384
Asn Glu Phe Val Val Ile His Asn Gly Ile Ile Thr Asn Tyr Lys Asp
          115          120          125

ctg agg aaa ttt ctg gaa agc aaa ggc tac gag ttt gag tca gaa aca      432
Leu Arg Lys Phe Leu Glu Ser Lys Gly Tyr Glu Phe Glu Ser Glu Thr
          130          135          140

gat aca gag acc atc gcc aag ctg att aaa tat gtg ttc gac aac aga      480
Asp Thr Glu Thr Ile Ala Lys Leu Ile Lys Tyr Val Phe Asp Asn Arg
145          150          155          160

gaa act gag gac att acg ttt tca acg ttg gtc gag aga gtc att cag      528
Glu Thr Glu Asp Ile Thr Phe Ser Thr Leu Val Glu Arg Val Ile Gln
          165          170          175

cag ttg gaa ggt gca ttc gcg ctg gtt ttc aag agt gtc cac tac cca      576
Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val His Tyr Pro
          180          185          190

gga gaa gcc gtt gcc aca cgg aga ggc agc ccc ctg ctc atc gga gtc      624
Gly Glu Ala Val Ala Thr Arg Arg Gly Ser Pro Leu Leu Ile Gly Val
          195          200          205

cgg agc aaa tac aag ctc tcc aca gaa cag atc cct atc tta tac agg      672
Arg Ser Lys Tyr Lys Leu Ser Thr Glu Gln Ile Pro Ile Leu Tyr Arg
          210          215          220

acg tgc act ctg gag aat gtg aag aat atc tgt aag aca cgg atg aag      720

```

Thr 225	Cys	Thr	Leu	Glu	Asn 230	Val	Lys	Asn	Ile	Cys 235	Lys	Thr	Arg	Met	Lys 240	
agg	ctg	gac	agc	tcc	gcc	tgc	ctg	cat	gct	gtg	ggc	gac	aag	gcc	gtg	768
Arg	Leu	Asp	Ser	Ser	Ala	Cys	Leu	His	Ala	Val	Gly	Asp	Lys	Ala	Val	
				245					250					255		
gaa	ttc	ttc	ttt	gct	tct	gat	gca	agc	gct	atc	ata	gag	cac	acc	aac	816
Glu	Phe	Phe	Phe	Ala	Ser	Asp	Ala	Ser	Ala	Ile	Ile	Glu	His	Thr	Asn	
			260					265					270			
cgg	gtc	atc	ttc	ctg	gag	gac	gat	gac	atc	gcc	gca	gtg	gct	gat	ggg	864
Arg	Val	Ile	Phe	Leu	Glu	Asp	Asp	Asp	Ile	Ala	Ala	Val	Ala	Asp	Gly	
			275				280					285				
aaa	ctc	tcc	att	cac	cgg	gtc	aag	cgc	tgc	gcc	agt	cat	cac	cat	cac	912
Lys	Leu	Ser	Ile	His	Arg	Val	Lys	Arg	Ser	Ala	Ser	His	His	His	His	
	290					295					300					
cat	cac	gat	gac	cca	tct	cga	gcc	atc	cag	acc	ttg	cag	atg	gaa	ctg	960
His	His	Asp	Asp	Pro	Ser	Arg	Ala	Ile	Gln	Thr	Leu	Gln	Met	Glu	Leu	
305					310				315					320		
cag	caa	atc	atg	aaa	ggc	aac	ttc	agt	gag	ttt	atg	cag	aag	gag	atc	1008
Gln	Gln	Ile	Met	Lys	Gly	Asn	Phe	Ser	Ala	Phe	Met	Gln	Lys	Glu	Ile	
				325					330					335		
ttc	gaa	cag	cca	gaa	tca	gtt	ttc	aat	act	atg	aga	ggc	cgg	gtg	aat	1056
Phe	Glu	Gln	Pro	Glu	Ser	Val	Phe	Asn	Thr	Met	Arg	Gly	Arg	Val	Asn	
			340					345					350			
ttt	gaa	acc	aac	aca	gtg	ctc	ctg	ggc	ggc	ttg	aag	gac	cac	ttg	aag	1104
Phe	Glu	Thr	Asn	Thr	Val	Leu	Leu	Gly	Gly	Leu	Lys	Asp	His	Leu	Lys	
			355				360					365				
gag	att	cga	cga	tgc	cga	cgg	ctc	atc	gtg	att	ggc	tgt	gga	acc	agc	1152
Glu	Ile	Arg	Arg	Cys	Arg	Arg	Leu	Ile	Val	Ile	Gly	Cys	Gly	Thr	Ser	
	370					375					380					
tac	cac	gct	gcc	gtg	gct	acg	cgg	caa	gtt	ttg	gag	gaa	ctg	act	gag	1200
Tyr	His	Ala	Ala	Val	Ala	Thr	Arg	Gln	Val	Leu	Glu	Glu	Leu	Thr	Glu	
385					390					395					400	
ctt	cct	gtg	atg	gtt	gaa	ctt	gct	agt	gat	ttt	ctg	gac	agg	aac	aca	1248
Leu	Pro	Val	Met	Val	Glu	Leu	Ala	Ser	Asp	Phe	Leu	Asp	Arg	Asn	Thr	
				405					410					415		
cct	gtg	ttc	agg	gat	gac	gtt	tgc	ttt	ttc	atc	agc	cag	tca	ggc	gag	1296
Pro	Val	Phe	Arg	Asp	Asp	Val	Cys	Phe	Phe	Ile	Ser	Gln	Ser	Gly	Glu	
			420				425						430			
acc	gag	gac	acc	ctc	ctg	gag	ctg	cgc	tac	tgt	aag	gac	cgc	ggc	gct	1344
Thr	Ala	Asp	Thr	Leu	Leu	Ala	Leu	Arg	Tyr	Cys	Lys	Asp	Arg	Gly	Ala	
			435				440					445				
ctc	acc	gtg	ggc	gtc	acc	aac	acc	gtg	ggc	agc	tcc	atc	tct	cgc	gag	1392
Leu	Thr	Val	Gly	Val	Thr	Asn	Thr	Val	Gly	Ser	Ser	Ile	Ser	Arg	Glu	
			450			455					460					
acc	gac	tgc	ggc	gtc	cac	atc	aac	gca	ggg	cgc	gag	gtc	ggc	gtg	gcc	1440
Thr	Asp	Cys	Gly	Val	His	Ile	Asn	Ala	Gly	Pro	Glu	Val	Gly	Val	Ala	
465					470				475						480	
agc	acc	aag	gct	tat	acc	agt	cag	ttc	atc	tct	ctg	gtg	atg	ttt	ggc	1488
Ser	Thr	Lys	Ala	Tyr	Thr	Ser	Gln	Phe	Ile	Ser	Leu	Val	Met	Phe	Gly	
			485					490							495	

ttg atg atg tct gaa gac cga att tca cta caa aac agg agg caa gag Leu Met Met Ser Glu Asp Arg Ile Ser Leu Gln Asn Arg Arg Gln Glu 500 505 510	1536
atc atc cgt ggc ttg aga tct tta cct gag ctg atc aag gaa gtg ctg Ile Ile Arg Gly Leu Arg Ser Leu Pro Glu Leu Ile Lys Glu Val Leu 515 520 525	1584
tct ctg gag gag aag atc cac gac ttg gcc ctg gag ctc tac acg cag Ser Leu Glu Glu Lys Ile His Asp Leu Ala Leu Glu Leu Tyr Thr Gln 530 535 540	1632
aga tct ctg ctg gtg atg ggg cgg ggc tac aac tat gcc acc tgc ctg Arg Ser Leu Leu Val Met Gly Arg Gly Tyr Asn Tyr Ala Thr Cys Leu 545 550 555 560	1680
gaa gga gcc ctg aaa att aaa gag ata acc tac atg cac tca gaa ggc Glu Gly Ala Leu Lys Ile Lys Glu Ile Thr Tyr Met His Ser Glu Gly 565 570 575	1728
atc ctg gct ggg gag ctg aag cac ggg ccc ctg gca ctg att gac aag Ile Leu Ala Gly Glu Leu Lys His Gly Pro Leu Ala Leu Ile Asp Lys 580 585 590	1776
cag atg ccc gtc atc atg gtc att atg aag gat cct tgc ttc gcc aaa Gln Met Pro Val Ile Met Val Ile Met Lys Asp Pro Cys Phe Ala Lys 595 600 605	1824
tgc cag aac gcc ctg cag caa gtc acg gcc cgc cag ggt cgc ccc att Cys Gln Asn Ala Leu Gln Gln Val Thr Ala Arg Gln Gly Arg Pro Ile 610 615 620	1872
ata ctg tgc tcc aag gac gat act gaa agt tcc aag ttt gcg tat aag Ile Leu Cys Ser Lys Asp Asp Thr Glu Ser Ser Lys Phe Ala Tyr Lys 625 630 635 640	1920
aca atc gag ctg ccc cac act gtg gac tgc ctc cag ggc atc ctg agc Thr Ile Glu Leu Pro His Thr Val Asp Cys Leu Gln Gly Ile Leu Ser 645 650 655	1968
gtg att ccg ctg cag ctg ctg tcc ttc cac ctg gct gtt ctc cga gga Val Ile Pro Leu Gln Leu Leu Ser Phe His Leu Ala Val Leu Arg Gly 660 665 670	2016
tat gac gtt gac ttc ccc aga aat ctg gcc aag tct gta act gtg gaa Tyr Asp Val Asp Phe Pro Arg Asn Leu Ala Lys Ser Val Thr Val Glu 675 680 685	2064
tga	2067
<210> 10	
<211> 688	
<212> PRT	
<213> Séquence artificielle	
<220>	
<223> GFAT2 modifiée par une étiquette de purification interne	
<400> 10	
Met Cys Gly Ile Phe Ala Tyr Met Asn Tyr Arg Val Pro Arg Thr Arg 1 5 10 15	

Lys Glu Ile Phe Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr  
 20 25 30  
 Arg Gly Tyr Asp Ser Ala Gly Val Ala Ile Asp Gly Asn Asn His Glu  
 35 40 45  
 Val Lys Glu Arg His Ile Gln Leu Val Lys Lys Arg Gly Lys Val Lys  
 50 55 60  
 Ala Leu Asp Glu Glu Leu Tyr Lys Gln Asp Ser Met Asp Leu Lys Val  
 65 70 75 80  
 Glu Phe Glu Thr His Phe Gly Ile Ala His Thr Arg Trp Ala Thr His  
 85 90 95  
 Gly Val Pro Ser Ala Val Asn Ser His Pro Gln Arg Ser Asp Lys Gly  
 100 105 110  
 Asn Glu Phe Val Val Ile His Asn Gly Ile Ile Thr Asn Tyr Lys Asp  
 115 120 125  
 Leu Arg Lys Phe Leu Glu Ser Lys Gly Tyr Glu Phe Glu Ser Glu Thr  
 130 135 140  
 Asp Thr Glu Thr Ile Ala Lys Leu Ile Lys Tyr Val Phe Asp Asn Arg  
 145 150 155 160  
 Glu Thr Glu Asp Ile Thr Phe Ser Thr Leu Val Glu Arg Val Ile Gln  
 165 170 175  
 Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val His Tyr Pro  
 180 185 190  
 Gly Glu Ala Val Ala Thr Arg Arg Gly Ser Pro Leu Leu Ile Gly Val  
 195 200 205  
 Arg Ser Lys Tyr Lys Leu Ser Thr Glu Gln Ile Pro Ile Leu Tyr Arg  
 210 215 220  
 Thr Cys Thr Leu Glu Asn Val Lys Asn Ile Cys Lys Thr Arg Met Lys  
 225 230 235 240  
 Arg Leu Asp Ser Ser Ala Cys Leu His Ala Val Gly Asp Lys Ala Val  
 245 250 255  
 Glu Phe Phe Phe Ala Ser Asp Ala Ser Ala Ile Ile Glu His Thr Asn  
 260 265 270  
 Arg Val Ile Phe Leu Glu Asp Asp Asp Ile Ala Ala Val Ala Asp Gly  
 275 280 285

Lys Leu Ser Ile His Arg Val Lys Arg Ser Ala Ser His His His His  
290 295 300

His His Asp Asp Pro Ser Arg Ala Ile Gln Thr Leu Gln Met Glu Leu  
305 310 315 320

Gln Gln Ile Met Lys Gly Asn Phe Ser Ala Phe Met Gln Lys Glu Ile  
325 330 335

Phe Glu Gln Pro Glu Ser Val Phe Asn Thr Met Arg Gly Arg Val Asn  
340 345 350

Phe Glu Thr Asn Thr Val Leu Leu Gly Gly Leu Lys Asp His Leu Lys  
355 360 365

Glu Ile Arg Arg Cys Arg Arg Leu Ile Val Ile Gly Cys Gly Thr Ser  
370 375 380

Tyr His Ala Ala Val Ala Thr Arg Gln Val Leu Glu Glu Leu Thr Glu  
385 390 395 400

Leu Pro Val Met Val Glu Leu Ala Ser Asp Phe Leu Asp Arg Asn Thr  
405 410 415

Pro Val Phe Arg Asp Asp Val Cys Phe Phe Ile Ser Gln Ser Gly Glu  
420 425 430

Thr Ala Asp Thr Leu Leu Ala Leu Arg Tyr Cys Lys Asp Arg Gly Ala  
435 440 445

Leu Thr Val Gly Val Thr Asn Thr Val Gly Ser Ser Ile Ser Arg Glu  
450 455 460

Thr Asp Cys Gly Val His Ile Asn Ala Gly Pro Glu Val Gly Val Ala  
465 470 475 480

Ser Thr Lys Ala Tyr Thr Ser Gln Phe Ile Ser Leu Val Met Phe Gly  
485 490 495

Leu Met Met Ser Glu Asp Arg Ile Ser Leu Gln Asn Arg Arg Gln Glu  
500 505 510

Ile Ile Arg Gly Leu Arg Ser Leu Pro Glu Leu Ile Lys Glu Val Leu  
515 520 525

Ser Leu Glu Glu Lys Ile His Asp Leu Ala Leu Glu Leu Tyr Thr Gln  
530 535 540

Arg Ser Leu Leu Val Met Gly Arg Gly Tyr Asn Tyr Ala Thr Cys Leu  
545 550 555 560

Glu Gly Ala Leu Lys Ile Lys Glu Ile Thr Tyr Met His Ser Glu Gly  
565 570 575

Ile Leu Ala Gly Glu Leu Lys His Gly Pro Leu Ala Leu Ile Asp Lys  
580 585 590

Gln Met Pro Val Ile Met Val Ile Met Lys Asp Pro Cys Phe Ala Lys  
595 600 605

Cys Gln Asn Ala Leu Gln Gln Val Thr Ala Arg Gln Gly Arg Pro Ile  
610 615 620

Ile Leu Cys Ser Lys Asp Asp Thr Glu Ser Ser Lys Phe Ala Tyr Lys  
625 630 635 640

Thr Ile Glu Leu Pro His Thr Val Asp Cys Leu Gln Gly Ile Leu Ser  
645 650 655

Val Ile Pro Leu Gln Leu Leu Ser Phe His Leu Ala Val Leu Arg Gly  
660 665 670

Tyr Asp Val Asp Phe Pro Arg Asn Leu Ala Lys Ser Val Thr Val Glu  
675 680 685

<210> 11  
<211> 2118  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> GFAT1Alt modifiée par une étiquette de purification interne

<220>  
<221> CDS  
<222> (1)..(2118)  
<223>

<220>  
<221> misc\_feature  
<222> (170)..(170)  
<223> t ou c

<400> 11  
atg tgt ggt ata ttt gct tac tta aac tac cat gtt cct cga acg aga 48  
Met Cys Gly Ile Phe Ala Tyr Leu Asn Tyr His Val Pro Arg Thr Arg  
1 5 10 15

cga gaa atc ctg gag acc cta atc aaa ggc ctt cag aga ctg gag tac 96  
Arg Glu Ile Leu Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr  
20 25 30

aga gga tat gat tct gct ggt gtg gga ttt gat gga ggc aat gat aaa 144  
Arg Gly Tyr Asp Ser Ala Gly Val Gly Phe Asp Gly Gly Asn Asp Lys

35					40					45						
gat	tgg	gaa	gcc	aat	gcc	tgc	aaa	anc	cag	ctt	att	aag	aag	aaa	gga	192
Asp	Trp	Glu	Ala	Asn	Ala	Cys	Lys	Xaa	Gln	Leu	Ile	Lys	Lys	Lys	Gly	
	50					55					60					
aaa	ggt	aag	gca	ctg	gat	gaa	gaa	ggt	cac	aag	caa	caa	gat	atg	gat	240
Lys	Val	Lys	Ala	Leu	Asp	Glu	Glu	Val	His	Lys	Gln	Gln	Asp	Met	Asp	
65					70					75					80	
ttg	gat	ata	gaa	ttt	gat	gta	cac	ctt	gga	ata	gct	cat	acc	cgt	tgg	288
Leu	Asp	Ile	Glu	Phe	Asp	Val	His	Leu	Gly	Ile	Ala	His	Thr	Arg	Trp	
				85					90					95		
gca	aca	cat	gga	gaa	ccc	agt	cct	gtc	aat	agc	cac	ccc	cag	cgc	tct	336
Ala	Thr	His	Gly	Glu	Pro	Ser	Pro	Val	Asn	Ser	His	Pro	Gln	Arg	Ser	
			100					105					110			
gat	aaa	aat	aat	gaa	ttt	atc	ggt	att	cac	aat	gga	atc	atc	acc	aac	384
Asp	Lys	Asn	Asn	Glu	Phe	Ile	Val	Ile	His	Asn	Gly	Ile	Ile	Thr	Asn	
		115					120					125				
tac	aaa	gac	ttg	aaa	aag	ttt	ttg	gaa	agc	aaa	ggc	tat	gac	ttc	gaa	432
Tyr	Lys	Asp	Leu	Lys	Lys	Phe	Leu	Glu	Ser	Lys	Gly	Tyr	Asp	Phe	Glu	
	130					135					140					
tct	gaa	aca	gac	aca	gag	aca	att	gcc	aag	ctc	ggt	aag	tat	atg	tat	480
Ser	Glu	Thr	Asp	Thr	Glu	Thr	Ile	Ala	Lys	Leu	Val	Lys	Tyr	Met	Tyr	
145					150					155				160		
gac	aat	cgg	gaa	agt	caa	gat	acc	agc	ttt	act	acc	ttg	gtg	gag	aga	528
Asp	Asn	Arg	Glu	Ser	Gln	Asp	Thr	Ser	Phe	Thr	Thr	Leu	Val	Glu	Arg	
				165					170					175		
ggt	atc	caa	caa	ttg	gaa	ggt	gct	ttt	gca	ctt	gtg	ttt	aaa	agt	ggt	576
Val	Ile	Gln	Gln	Leu	Glu	Gly	Ala	Phe	Ala	Leu	Val	Phe	Lys	Ser	Val	
			180					185					190			
cat	ttt	ccc	ggg	caa	gca	ggt	ggc	aca	agg	cga	ggt	agc	cct	ctg	ttg	624
His	Phe	Pro	Gly	Gln	Ala	Val	Gly	Thr	Arg	Arg	Gly	Ser	Pro	Leu	Leu	
		195					200					205				
att	ggt	gta	cgg	agt	gaa	cat	aaa	ctt	tct	act	gat	cac	att	cct	ata	672
Ile	Gly	Val	Arg	Ser	Glu	His	Lys	Leu	Ser	Thr	Asp	His	Ile	Pro	Ile	
	210					215					220					
ctc	tac	aga	aca	gct	agg	act	cag	att	gga	tca	aaa	ttc	aca	cgg	tgg	720
Leu	Tyr	Arg	Thr	Ala	Arg	Thr	Gln	Ile	Gly	Ser	Lys	Phe	Thr	Arg	Trp	
225					230				235						240	
gga	tca	cag	gga	gaa	aga	ggc	aaa	gac	aag	aaa	gga	agc	tgc	aat	ctc	768
Gly	Ser	Gln	Gly	Glu	Arg	Gly	Lys	Asp	Lys	Lys	Gly	Ser	Cys	Asn	Leu	
				245					250					255		
tct	cgt	gtg	gac	agc	aca	acc	tgc	ctt	ttc	ccg	gtg	gaa	gaa	aaa	gca	816
Ser	Arg	Val	Asp	Ser	Thr	Thr	Cys	Leu	Phe	Pro	Val	Glu	Glu	Lys	Ala	
			260					265					270			
gtg	gag	tat	tac	ttt	gct	tct	gat	gca	agt	gct	gtc	ata	gaa	cac	acc	864
Val	Glu	Tyr	Tyr	Phe	Ala	Ser	Asp	Ala	Ser	Ala	Val	Ile	Glu	His	Thr	
		275					280					285				
aat	cgc	gtc	atc	ttt	ctg	gaa	gat	gat	gat	gtt	gca	gca	gta	gtg	gat	912
Asn	Arg	Val	Ile	Phe	Leu	Glu	Asp	Asp	Asp	Val	Ala	Ala	Val	Val	Asp	
	290					295					300					



gga cgt ctt tct atc cat cga att aaa cga act gca gga cat cac cat Gly Arg Leu Ser Ile His Arg Ile Lys Arg Thr Ala Gly His His His 305 310 315 320	960
cac cat cac gat cac ccc gga cga gct gtg caa aca ctc cag atg gaa His His His Asp His Pro Gly Arg Ala Val Gln Thr Leu Gln Met Glu 325 330 335	1008
ctc cag cag atc atg aag ggc aac ttc agt tca ttt atg cag aag gaa Leu Gln Gln Ile Met Lys Gly Asn Phe Ser Ser Phe Met Gln Lys Glu 340 345 350	1056
ata ttt gag cag cca gag tct gtc gtg aac aca atg aga gga aga gtc Ile Phe Glu Gln Pro Glu Ser Val Val Asn Thr Met Arg Gly Arg Val 355 360 365	1104
aac ttt gat gac tat act gtg aat ttg ggt ggt ttg aag gat cac ata Asn Phe Asp Asp Tyr Thr Val Asn Leu Gly Gly Leu Lys Asp His Ile 370 375 380	1152
aag gag atc cag aga tgc cgg cgt ttg att ctt att gct tgt gga aca Lys Glu Ile Gln Arg Cys Arg Arg Leu Ile Leu Ile Ala Cys Gly Thr 385 390 400	1200
agt tac cat gct ggt gta gca aca cgt caa gtt ctt gag gag ctg act Ser Tyr His Ala Gly Val Ala Thr Arg Gln Val Leu Glu Glu Leu Thr 405 410 415	1248
gag ttg cct gtg atg gtg gaa cta gca agt gac ttc ctg gac aga aac Glu Leu Pro Val Met Val Glu Leu Ala Ser Asp Phe Leu Asp Arg Asn 420 425 430	1296
aca cca gtc ttt cga gat gat gtt tgc ttt ttc ctt agt caa tca ggt Thr Pro Val Phe Arg Asp Asp Val Cys Phe Phe Leu Ser Gln Ser Gly 435 440 445	1344
gag aca gca gat act ttg atg ggt ctt cgt tac tgt aag gag aga gga Glu Thr Ala Asp Thr Leu Met Gly Leu Arg Tyr Cys Lys Glu Arg Gly 450 455 460	1392
gct tta act gtg ggg atc aca aac aca gtt ggc agt tcc ata tca cgg Ala Leu Thr Val Gly Ile Thr Asn Thr Val Gly Ser Ser Ile Ser Arg 465 470 475 480	1440
gag aca gat tgt gga gtt cat att aat gct ggt cct gag att ggt gtg Glu Thr Asp Cys Gly Val His Ile Asn Ala Gly Pro Glu Ile Gly Val 485 490 495	1488
gcc agt aca aag gct tat acc agc cag ttt gta tcc ctt gtg atg ttt Ala Ser Thr Lys Ala Tyr Thr Ser Gln Phe Val Ser Leu Val Met Phe 500 505 510	1536
gcc ctt atg atg tgt gat gat cgg atc tcc atg caa gaa aga cgc aaa Ala Leu Met Met Cys Asp Asp Arg Ile Ser Met Gln Glu Arg Arg Lys 515 520 525	1584
gag atc atg ctt gga ttg aaa cgg ctg cct gat ttg att aag gaa gta Glu Ile Met Leu Gly Leu Lys Arg Leu Pro Asp Leu Ile Lys Glu Val 530 535 540	1632
ctg agc atg gat gac gaa att cag aaa cta gca aca gaa ctt tat cat Leu Ser Met Asp Asp Glu Ile Gln Lys Leu Ala Thr Glu Leu Tyr His 545 550 555 560	1680
cag aag tca gtt ctg ata atg gga cga ggc tat cat tat gct act tgt Gln Lys Ser Val Leu Ile Met Gly Arg Gly Tyr His Tyr Ala Thr Cys	1728

565					570					575					
ctt gaa ggg gca	ctg aaa atc aaa	gaa att act tat	atg cac tct gaa											1776	
Leu Glu Gly Ala	Leu Lys Ile Lys	Glu Ile Thr Tyr	Met His Ser Glu	580	585					590					
ggc atc ctt gct	ggt gaa ttg aaa	cat ggc cct ctg	gct ttg gtg gat											1824	
Gly Ile Leu Ala	Gly Glu Leu Lys	His Gly Pro Leu	Ala Leu Val Asp	595	600					605					
aaa ttg atg cct	gtg atc atg atc	atg aga gat cac	act tat gcc											1872	
Lys Leu Met Pro	Val Ile Met Ile	Ile Met Arg Asp	His Thr Tyr Ala	610	615					620					
aag tgt cag aat	gct ctt cag caa	gtg gtt gct cgg	cag ggg cgg cct											1920	
Lys Cys Gln Asn	Ala Leu Gln Gln	Val Val Ala Arg	Gln Gly Arg Pro	625	630					635				640	
gtg gta att tgt	gat aag gag gat	act gag acc att	aag aac aca aaa											1968	
Val Val Ile Cys	Asp Lys Glu Asp	Thr Glu Thr Ile	Lys Asn Thr Lys	645						650				655	
aga acg atc aag	gtg ccc cac tca	gtg gac tgc ttg	cag ggc att ctc											2016	
Arg Thr Ile Lys	Val Pro His Ser	Val Asp Cys Leu	Gln Gly Ile Leu	660						665				670	
agc gtg atc cct	tta cag ttg ctg	gct ttc cac ctt	gct gtg ctg aga											2064	
Ser Val Ile Pro	Leu Gln Leu Leu	Ala Phe His Leu	Ala Val Leu Arg	675						680				685	
ggc tat gat gtt	gat ttc cca cgg	aat ctt gcc aaa	tct gtg act gta											2112	
Gly Tyr Asp Val	Asp Phe Pro Arg	Asn Leu Ala Lys	Ser Val Thr Val	690	695					700					
gag tga														2118	
Glu															
705															

<210> 12  
 <211> 705  
 <212> PRT  
 <213> Séquence artificielle

<220>  
 <221> misc\_feature  
 <222> (57)..(57)  
 <223> Le 'Xaa' en position 57 représente Thr ou Ile.

<220>  
 <223> GFAT1Alt modifiée par une étiquette de purification interne

<400> 12

Met Cys Gly Ile	Phe Ala Tyr Leu	Asn Tyr His Val	Pro Arg Thr Arg
1	5	10	15

Arg Glu Ile Leu	Glu Thr Leu Ile	Lys Gly Leu Gln	Arg Leu Glu Tyr
20	25	30	

Arg Gly Tyr Asp	Ser Ala Gly Val	Gly Phe Asp Gly	Gly Asn Asp Lys
35	40	45	

Asp Trp Glu Ala Asn Ala Cys Lys Xaa Gln Leu Ile Lys Lys Lys Gly  
 50 55 60  
 Lys Val Lys Ala Leu Asp Glu Glu Val His Lys Gln Gln Asp Met Asp  
 65 70 75 80  
 Leu Asp Ile Glu Phe Asp Val His Leu Gly Ile Ala His Thr Arg Trp  
 85 90 95  
 Ala Thr His Gly Glu Pro Ser Pro Val Asn Ser His Pro Gln Arg Ser  
 100 105 110  
 Asp Lys Asn Asn Glu Phe Ile Val Ile His Asn Gly Ile Ile Thr Asn  
 115 120 125  
 Tyr Lys Asp Leu Lys Lys Phe Leu Glu Ser Lys Gly Tyr Asp Phe Glu  
 130 135 140  
 Ser Glu Thr Asp Thr Glu Thr Ile Ala Lys Leu Val Lys Tyr Met Tyr  
 145 150 155 160  
 Asp Asn Arg Glu Ser Gln Asp Thr Ser Phe Thr Thr Leu Val Glu Arg  
 165 170 175  
 Val Ile Gln Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val  
 180 185 190  
 His Phe Pro Gly Gln Ala Val Gly Thr Arg Arg Gly Ser Pro Leu Leu  
 195 200 205  
 Ile Gly Val Arg Ser Glu His Lys Leu Ser Thr Asp His Ile Pro Ile  
 210 215 220  
 Leu Tyr Arg Thr Ala Arg Thr Gln Ile Gly Ser Lys Phe Thr Arg Trp  
 225 230 235 240  
 Gly Ser Gln Gly Glu Arg Gly Lys Asp Lys Lys Gly Ser Cys Asn Leu  
 245 250 255  
 Ser Arg Val Asp Ser Thr Thr Cys Leu Phe Pro Val Glu Glu Lys Ala  
 260 265 270  
 Val Glu Tyr Tyr Phe Ala Ser Asp Ala Ser Ala Val Ile Glu His Thr  
 275 280 285  
 Asn Arg Val Ile Phe Leu Glu Asp Asp Asp Val Ala Ala Val Val Asp  
 290 295 300  
 Gly Arg Leu Ser Ile His Arg Ile Lys Arg Thr Ala Gly His His His  
 305 310 315 320

His His His Asp His Pro Gly Arg Ala Val Gln Thr Leu Gln Met Glu  
325 330 335

Leu Gln Gln Ile Met Lys Gly Asn Phe Ser Ser Phe Met Gln Lys Glu  
340 345 350

Ile Phe Glu Gln Pro Glu Ser Val Val Asn Thr Met Arg Gly Arg Val  
355 360 365

Asn Phe Asp Asp Tyr Thr Val Asn Leu Gly Gly Leu Lys Asp His Ile  
370 375 380

Lys Glu Ile Gln Arg Cys Arg Arg Leu Ile Leu Ile Ala Cys Gly Thr  
385 390 395 400

Ser Tyr His Ala Gly Val Ala Thr Arg Gln Val Leu Glu Glu Leu Thr  
405 410 415

Glu Leu Pro Val Met Val Glu Leu Ala Ser Asp Phe Leu Asp Arg Asn  
420 425 430

Thr Pro Val Phe Arg Asp Asp Val Cys Phe Phe Leu Ser Gln Ser Gly  
435 440 445

Glu Thr Ala Asp Thr Leu Met Gly Leu Arg Tyr Cys Lys Glu Arg Gly  
450 455 460

Ala Leu Thr Val Gly Ile Thr Asn Thr Val Gly Ser Ser Ile Ser Arg  
465 470 475 480

Glu Thr Asp Cys Gly Val His Ile Asn Ala Gly Pro Glu Ile Gly Val  
485 490 495

Ala Ser Thr Lys Ala Tyr Thr Ser Gln Phe Val Ser Leu Val Met Phe  
500 505 510

Ala Leu Met Met Cys Asp Asp Arg Ile Ser Met Gln Glu Arg Arg Lys  
515 520 525

Glu Ile Met Leu Gly Leu Lys Arg Leu Pro Asp Leu Ile Lys Glu Val  
530 535 540

Leu Ser Met Asp Asp Glu Ile Gln Lys Leu Ala Thr Glu Leu Tyr His  
545 550 555 560

Gln Lys Ser Val Leu Ile Met Gly Arg Gly Tyr His Tyr Ala Thr Cys  
565 570 575

Leu Glu Gly Ala Leu Lys Ile Lys Glu Ile Thr Tyr Met His Ser Glu  
580 585 590

Gly Ile Leu Ala Gly Glu Leu Lys His Gly Pro Leu Ala Leu Val Asp  
595 600 605

Lys Leu Met Pro Val Ile Met Ile Ile Met Arg Asp His Thr Tyr Ala  
610 615 620

Lys Cys Gln Asn Ala Leu Gln Gln Val Val Ala Arg Gln Gly Arg Pro  
625 630 635 640

Val Val Ile Cys Asp Lys Glu Asp Thr Glu Thr Ile Lys Asn Thr Lys  
645 650 655

Arg Thr Ile Lys Val Pro His Ser Val Asp Cys Leu Gln Gly Ile Leu  
660 665 670

Ser Val Ile Pro Leu Gln Leu Leu Ala Phe His Leu Ala Val Leu Arg  
675 680 685

Gly Tyr Asp Val Asp Phe Pro Arg Asn Leu Ala Lys Ser Val Thr Val  
690 695 700

Glu  
705

<210> 13  
<211> 608  
<212> PRT  
<213> Escherichia coli

<400> 13

Cys Gly Ile Val Gly Ala Ile Ala Gln Arg Asp Val Ala Glu Ile Leu  
1 5 10 15

Leu Glu Gly Leu Arg Arg Leu Glu Tyr Arg Gly Tyr Asp Ser Ala Gly  
20 25 30

Leu Ala Val Val Asp Ala Glu Gly His Met Thr Arg Leu Arg Arg Leu  
35 40 45

Gly Lys Val Gln Met Leu Ala Gln Ala Ala Glu Glu His Pro Leu His  
50 55 60

Gly Gly Thr Gly Ile Ala His Thr Arg Trp Ala Thr His Gly Glu Pro  
65 70 75 80

Ser Glu Val Asn Ala His Pro His Val Ser Glu His Ile Val Val Val  
85 90 95

His Asn Gly Ile Ile Glu Asn His Glu Pro Leu Arg Glu Glu Leu Lys  
100 105 110

Ala Arg Gly Tyr Thr Phe Val Ser Glu Thr Asp Thr Glu Val Ile Ala  
115 120 125

His Leu Val Asn Trp Glu Leu Lys Gln Gly Gly Thr Leu Arg Glu Ala  
130 135 140

Val Leu Arg Ala Ile Pro Gln Leu Arg Gly Ala Tyr Gly Thr Val Ile  
145 150 155 160

Met Asp Ser Arg His Pro Asp Thr Leu Leu Ala Ala Arg Ser Gly Ser  
165 170 175

Pro Leu Val Ile Gly Leu Gly Met Gly Glu Asn Phe Ile Ala Ser Asp  
180 185 190

Gln Leu Ala Leu Leu Pro Val Thr Arg Arg Phe Ile Phe Leu Glu Glu  
195 200 205

Gly Asp Ile Ala Glu Ile Thr Arg Arg Ser Val Asn Ile Phe Asp Lys  
210 215 220

Thr Gly Ala Glu Val Lys Arg Gln Asp Ile Glu Ser Asn Leu Gln Tyr  
225 230 235 240

Asp Ala Gly Asp Lys Gly Ile Tyr Arg His Tyr Met Gln Lys Glu Ile  
245 250 255

Tyr Glu Gln Pro Asn Ala Ile Lys Asn Thr Leu Thr Gly Arg Ile Ser  
260 265 270

His Gly Gln Val Asp Leu Ser Glu Leu Gly Pro Asn Ala Asp Glu Leu  
275 280 285

Leu Ser Lys Val Glu His Ile Gln Ile Leu Ala Cys Gly Thr Ser Tyr  
290 295 300

Asn Ser Gly Met Val Ser Arg Tyr Trp Phe Glu Ser Leu Ala Gly Ile  
305 310 315 320

Pro Cys Asp Val Glu Ile Ala Ser Glu Phe Arg Tyr Arg Lys Ser Ala  
325 330 335

Val Arg Arg Asn Ser Leu Met Ile Thr Leu Ser Gln Ser Gly Glu Thr  
340 345 350

Ala Asp Thr Leu Ala Gly Leu Arg Leu Ser Lys Glu Leu Gly Tyr Leu

355

360

365

Gly Ser Leu Ala Ile Cys Asn Val Pro Gly Ser Ser Leu Val Arg Glu  
370 375 380

Ser Asp Leu Ala Leu Met Thr Asn Ala Gly Thr Glu Ile Gly Val Ala  
385 390 395 400

Ser Thr Lys Ala Phe Thr Thr Gln Leu Thr Val Leu Leu Met Leu Val  
405 410 415

Ala Lys Leu Ser Arg Leu Lys Gly Leu Asp Ala Ser Ile Glu His Asp  
420 425 430

Ile Val His Gly Leu Gln Ala Leu Pro Ser Arg Ile Glu Gln Met Leu  
435 440 445

Ser Gln Asp Lys Arg Ile Glu Ala Leu Ala Glu Asp Phe Ser Asp Lys  
450 455 460

His His Ala Leu Phe Leu Gly Arg Gly Asp Gln Tyr Pro Ile Ala Leu  
465 470 475 480

Glu Gly Ala Leu Lys Leu Lys Glu Ile Ser Tyr Ile His Ala Glu Ala  
485 490 495

Tyr Ala Ala Gly Glu Leu Lys His Gly Pro Leu Ala Leu Ile Asp Ala  
500 505 510

Asp Met Pro Val Ile Val Val Ala Pro Asn Asn Glu Leu Leu Glu Lys  
515 520 525

Leu Lys Ser Asn Ile Glu Glu Val Arg Ala Arg Gly Gly Gln Leu Tyr  
530 535 540

Val Phe Ala Asp Gln Asp Ala Gly Phe Val Ser Ser Asp Asn Met His  
545 550 555 560

Ile Ile Glu Met Pro His Val Glu Glu Val Ile Ala Pro Ile Phe Tyr  
565 570 575

Thr Val Pro Leu Gln Leu Leu Ala Tyr His Val Ala Leu Ile Lys Gly  
580 585 590

Thr Asp Val Asp Gln Pro Arg Asn Leu Ala Lys Ser Val Thr Val Glu  
595 600 605

<210> 14  
<211> 72  
<212> ADN

<213> Séquence artificielle

<220>

<223> Amorce

<400> 14

tggacgtctt tctatccatc gaattaaacg aactgcagga catcaccatc accatcacga 60  
tcaccccgga cg 72

<210> 15

<211> 44

<212> ADN

<213> Séquence artificielle

<220>

<223> Amorce

<400> 15

caaagttgac tcttcctctc attgtgttca cgacagactc tggc 44

<210> 16

<211> 43

<212> ADN

<213> Séquence artificielle

<220>

<223> Amorce

<400> 16

aatctagatt catgctcgag cggccgccag tgtgattgat atc 43

<210> 17

<211> 36

<212> ADN

<213> Séquence artificielle

<220>

<223> Amorce

<400> 17

atttttatca gagcgctggg ggtggctatt gacagg 36

<210> 18

<211> 8

<212> PRT

<213> Séquence artificielle

<220>

<223> Etiquette FLAG

<400> 18

Asp Thr Lys Asp Asp Asp Asp Lys  
1 5

<210> 19

<211> 6

<212> PRT

<213> Séquence artificielle



<220>

<223> Etiquette hexa-histidine

<400> 19

His His His His His His  
1 5



# BREVET D'INVENTION

## CERTIFICAT D'UTILITÉ

Code de la propriété intellectuelle - Livre VI



N° 11235\*0

## DÉPARTEMENT DES BREVETS


26 bis, rue de Saint Pétersbourg  
75800 Paris Cedex 08

Téléphone : 01 53 04 53 04 Télécopie : 01 42 94 86 54

DÉSIGNATION D'INVENTEUR(S) Page N° 1.1 / 1.1  
(Si le demandeur n'est pas l'inventeur ou l'unique inventeur)

Cet imprimé est à remplir lisiblement à l'encre noire

DB 113 W / 2608

<b>Vos références pour ce dossier</b> (facultatif)		IFB 03 BP CNR GFAT	
<b>N° D'ENREGISTREMENT NATIONAL</b>		03/08350	
<b>TITRE DE L'INVENTION</b> (200 caractères ou espaces maximum)			
GLUTAMINE :FRUCTOSE-6-PHASPHATE AMIDOTRANSFERASE (GFAT) COMPRENANT UNE ETIQUETTE DE PURIFICATION INTERNE, ET SON UTILISATION POUR LE CRIBLAGE DE COMPOSES			
<b>LE(S) DEMANDEUR(S) :</b>			
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE 3, rue Michel-Ange F-75794 PARIS CEDEX 16, France			
<b>DESIGNE(NT) EN TANT QU'INVENTEUR(S) :</b> (Indiquez en haut à droite «Page N° 1/1» S'il y a plus de trois inventeurs, utilisez un formulaire identique et numérotez chaque page en indiquant le nombre total de pages).			
Nom		BADET-DENISOT	
Prénoms		Marie-Ange, Juliette, Etiennette	
Adresse	Rue	2, rue d'Adélaïau	
	Code postal et ville	91470 FORGES-LES BAINS	
Société d'appartenance (facultatif)			
Nom		BADET	
Prénoms		Bernard, François	
Adresse	Rue	2, rue d'Adélaïau	
	Code postal et ville	91470 FORGES-LES BAINS	
Société d'appartenance (facultatif)			
Nom			
Prénoms			
Adresse	Rue		
	Code postal et ville		
Société d'appartenance (facultatif)			
<b>DATE ET SIGNATURE(S)</b> <b>DU (DES) DEMANDEUR(S)</b> <b>OU DU MANDATAIRE</b> (Nom et qualité du signataire)		Paris, le 9 septembre 2003  Charles DEMACHY, Mandataire 422.5/PP170 	

**PCT/FR2004/001800**



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ **BLACK BORDERS**

☒ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☒ **FADED TEXT OR DRAWING**

☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**